System Level Analysis of LTE-Advanced

with Emphasis on Multi-Component Carrier Management

Wang, Yuanye

Publication date:
2010

Document Version
Publisher's PDF, also known as Version of record

System Level Analysis of LTE-Advanced:
with Emphasis on Multi-Component Carrier Management

PhD Thesis

by

Yuanye Wang

A dissertation submitted to
Department of Electronic Systems,
the Faculty of Engineering and Science, Aalborg University
in partial fulfillment for the degree of
Doctor of Philosophy,
Aalborg, Denmark
September 2010.

Supervisors:
Preben E. Mogensen, PhD,
Professor, Aalborg University, Denmark.
Troels B. Sørensen, PhD,
Associate Professor, Aalborg University, Denmark.
Klaus I. Pedersen, PhD,
Senior Wireless Network Specialist, Nokia Siemens Networks, Aalborg, Denmark.

Opponents:
Petar Popovski, PhD
Associate Professor, Aalborg University, Denmark.
David Astély, PhD,
Technical Coordinator TDD LTE Research, Ericsson Research, Kista, Sweden.
Raymond Knopp, PhD,
Professor, EURECOM, France.

ISBN: 978-87-92328-56-4

Abstract

This PhD thesis focuses on system level analysis of Multi-Component Carrier (CC) management for Long Term Evolution (LTE)-Advanced. Cases where multiple CCs are aggregated to form a larger bandwidth are studied. The analysis is performed for both local area and wide area networks.

In local area, Time Division Duplexing (TDD) is chosen as the duplexing mode in this study. The performance with different network time synchronization levels is compared, and it is observed that achieving time synchronization significantly improves the uplink performance without penalizing much of the downlink transmission. Next the technique of frequency reuse is investigated. As compared to reuse-1, using different frequency channels in neighboring cells reduces the interference to offer large performance gain. To avoid the frequency planning, several decentralized algorithms are developed for interference reduction. Compared to the case of reuse-1, they achieve a gain of 50∼500% in cell edge user throughput, with small or no loss in average cell throughput.

For the wide area network, effort is devoted to the downlink of LTE-Advanced. Such a system is assumed to be backwards compatible to LTE release 8, i.e., some users can access all CCs (LTE-Advanced users), whereas some are restricted to operate within a single CC (release 8 users). First, load balancing across the multiple CCs is analyzed. Several known approaches are studied and the best one is identified. A cross-CC packet scheduler is afterwards proposed. It improves the cell edge user throughput by up to 90% over the independent scheduling with full buffer transmission and 40% with finite buffer transmission, depending primarily on the ratio of LTE-Advanced users. Meanwhile, there is no loss in the average cell throughput.

The channel aware packet scheduling and link adaptation require feedback of Chan-

nel Quality Indicator (CQI) and acknowledgement of packet receptions (ACK/NACKs) across the CCs. This gives rise to potentially high uplink overhead. Reduction of the feedback overhead is therefore investigated. A load adaptive CQI reduction scheme is recommended. It reduces the CQI by 94% at low load, and 79∼93% at medium to high load, with reasonable loss in downlink performance. To reduce the ACK/NACK feedback, multiple ACK/NACKs can be bundled, with slightly degraded downlink throughput.

Dansk Resumé[1]

Denne PhD afhandling har fokus på Multi-Component Carrier (CC) Management for "Long Term Evolution Advanced - LTE-Advanced". Som en del af dette studeres tilfælde med aggregering af flere frekvenser båndbredde for at opnå en større total båndbredde.

Time division duplexing (TDD) er antaget i dette studium for små lokale basisstationer. Forskellige grader af synkronisering er blevet studeret, og det konkluderes, at præcis tidssynkronisering er vigtig for god performance. Forskellige frekvensgenbrug er også blevet undersøgt. I forhold til frekvens genbrug 1:1, viser det sig, at brug af forskellige frevenser i naboceller reducerer interferensen og giver betydelig bedre performance. For at undgå manuel frekvensplanlægning er der blevet udviklet flere forskellige decentrale algoritmer, som klarer dette automatisk. Resultaterne viser, at brug at de udviklede decentrale algoritmer giver 50-100% bedre performance i forhold til brug af frekvensgenbrug 1:1.

For macrocelle netværk har fokus været på studier af Radio Resource Management for konfigurationer med aggregering af adskillige frekvens båndbredder. Tilfælde, hvor nogle brugere kun kan bruge én frekvens, mens andre kan få allokeret flere frekvenser på samme tid (aggregering), undersøges. Load balancing algoritmer til sådanne systemer er blevet analyseret. En simpel, men effektiv pakke skeduler er herefter blevet udviklet. Denne algoritme kan forbedre performance med helt op til 90% i forhold til tilfælde med standard uafhængig pakkeskedulering per frekvensbåndbredde.

Radiokanal afhængig pakke skedulering og link adaptation kræver feedback til basisstationen. Dette kan i værste fald give anledning til stort signaleringsoverhead.

[1] Translated by Klaus I. Pedersen of Nokia Siemens Networks, Aalborg, Denmark.

Forskellige teknikker til reduktion af signallerings-overhead er derfor blevet studeret. En såkaldt "load adaptive" løsning anbefaldes. Med sådanne teknikker kan signallerings overhead reduceres med helt op til 94% for tilfælde med lav traffik belastning, og 79-93% for tilfælde med høj traffik belastning. For yderligere reduktion af signallerings overhead, kan "bundling of ACK/NACKs" også benyttes, med kun et lille tab af downlink performance.

Preface and Acknowledgments

This PhD thesis is the result of a three-year research project carried out at the Radio Access Technology (RATE) section, Institute of Electronic Systems, Aalborg University, Denmark. The thesis work has been completed in parallel with the mandatory courses and teaching/working obligations required in order to obtain the PhD degree. The research project has been completed under the supervision and guidance of Professor Preben E. Mogensen (Aalborg University, Nokia Siemens Networks, Aalborg, Denmark), Associate Professor Troels B. Sørensen (Aalborg University) and Senior Wireless Network Specialist Dr. Klaus I. Pedersen (Nokia Siemens Networks, Aalborg, Denmark). It has been co-financed by the Faculty of Engineering, Science and Medicine, Aalborg University (1/3), Nokia Siemens Networks, Aalborg (1/3) and the Danish Agency for Science, Technology and Innovation - Forskeruddannelsesudvalget (FUU) (1/3).

The thesis investigates system level technologies to improve the performance of the Long Term Evolution (LTE)-Advanced systems, with emphasis on multi-Component Carrier management. The chapters in the thesis fall into two parts, where the first part deals with the local area network performance, and the second part addresses the wide area networks. The reader is expected to have a basic knowledge of system level aspects of LTE.

First of all, I would like to express my gratitude to the supervisors of my master thesis, Dr. Muhammad Imadur Rahman, Dr. Suvra Das and Associate Professor Troels B. Sørensen for introducing me to the main research area of this study. I am also sincerely grateful to my PhD supervisors for their technical supports and guidance throughout the study. They have been very patient in the numerous discussions and have offered me a lot of good advices. Also, their help in reviewing/correcting my work is highly appreciated.

I would like to thank Dr. Muhammad Imadur Rahman, Dr. Simone Frattasi and Dr. Nicola Marchetti, who guided me in formulating the work area in the beginning of the study. Without their help, I could not have completed this work. I am also thankful to my colleagues in the RATE group and Nokia Siemens Networks, Aalborg. In particular, thanks to Mads Brix and Jens Steiner for their help in simulator programming; Dr. István Kovács for the prompt and accurate answers regarding the LTE network setups; Daniela Laselva for the explanation of many related algorithms; Frank Frederiksen and Dr. Troels Kolding for the IT support.

I am also deeply thankful to our section secretary Lisbeth Schiønning Larsen for the friendly assistance with administrative issues, and Jytte Larsen, the secretary of Nokia Siemens Networks (Aalborg site), for the proofreading of my papers and thesis.

Finally, special thanks to my girlfriend Zhen Liu, who has offered me great help both at work and after work.

Yuanye Wang, Sep 2010

Contents

Chapter 1

Introduction

In Section 1.1, the background of the mobile communication systems related with this PhD study is given; Section 1.2 gives a short introduction of Long Term Evolution (LTE)-Advanced; Section 1.3 provides the state of art of the relevant topics; The problem of the study is defined in Section 1.4 and the research methodology is described in Section 1.5; Section 1.6 summarizes the contributions in the thesis and Section 1.7 provides an outline of the chapters and appendices.

1.1 Preliminaries

Mobile networks have experienced dramatic growth during the past decades. Other than voice transmission, the current mobile networks can provide users with a variety of services, including web browsing, video streaming, file downloading etc. The capability of high data rate transmission is one of the most important enablers for the wide application of wireless networks. It ultimately determines the kind of service that can be provided to the users and the service quality each user will experience.

To accommodate the increasing amount of traffic requirement, the data rate of a wireless network has increased, along with the bandwidth extension. The First Generation (1G) Advanced Mobile Phone System (AMPS) system [1], which was developed in the late 1970s and first deployed in 1983 [2], uses a narrow channel of

30 kHz to offer data rate of 10 kbps. The Second Generation (2G) Global System for Mobile communication (GSM) system occupies a much wider channel of 200 kHz, which is shared by 8 users in time domain [3]. While the voice transmission has a similar data rate as AMPS, GSM offers an data rate of several tens of kilobits per second for the data traffic [4]. Wideband Code Division Multiple Access (WCDMA), which is known as the Third Generation (3G) wireless system, was first deployed in 2001. It offers maximum 2 Mbps to each user in the downlink and 768 kbps in the uplink with 5 MHz bandwidth [5–7]. Recently, the Worldwide Interoperability for Microwave Access (WiMAX) and LTE have also been commercialized [8–11]. With up to 20 MHz bandwidth, LTE provides users with peak data of 300 Mbps in the downlink, and 150 Mbps in the uplink [9–11]. WiMAX also supports high data rate of tens to hundreds of megabits per second to the users [8]. Along with the increasing data rate, the network has also changed from mainly circuit switched (1G and 2G), to a mixture of circuit and packet switched (WCDMA), and purely packet switched (LTE, WiMAX) [11, 12].

With the roll-out of 3G in many countries, researchers and standardization organizations have increased their efforts to offer even higher data rate and more services to the users than 3G. To regulate the research activities, the International Telecommunications Union - Radio Communication Sector (ITU-R) has defined the concept and requirements of the International Mobile Telecommunications (IMT)-Advanced [13, 14]. According to IMT-Advanced, the downlink peak data rate should be at least 1 Gbps; and most importantly, high data rates should be provided over a large portion of the cell. There are many possible ways to satisfy the IMT-Advanced requirements, e.g., by evolving LTE [15] or WiMAX systems [16,17]. This PhD work will focus on system level analysis of evolving the LTE system towards the IMT-Advanced requirements. This evolved LTE system is generally recognized as LTE-Advanced. The aim is to provide recommendations for both traditional wide area networks and the emerging local area networks.

1.2 Outline of LTE-Advanced

This section gives an overview of LTE-Advanced. Some relevant topics for this thesis are addressed, including bandwidth extension, Radio Resource Management (RRM), feedback report, duplexing, and the target scenarios.

1.2.1 Bandwidth Extension Using Carrier Aggregation

To achieve higher peak data rate, the bandwidth of the LTE-Advanced systems should be much wider than current 3G systems. It is expected to be as high as 100 MHz [14–16]. Several carriers can be aggregated to form a wide bandwidth. This is

known as carrier aggregation and the aggregated carriers are named as Component Carrier (CC)s in Third Generation Partnership Project (3GPP) [18–20]. Fig. 1.1 gives an example where 5 CCs, each with 20 MHz bandwidth, are aggregated to form a 100 MHz bandwidth.

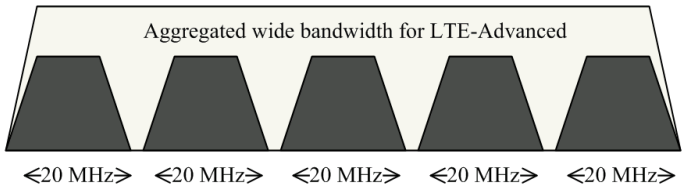


Fig. 1.1: Carrier aggregation of 5 CCs to form a 100 MHz bandwidth.

A general introduction about carrier aggregation and related technical challenges is provided in [19, 20]. Different aggregation types are outlined, e.g., the contiguous and non-contiguous carrier aggregations. In [18], the requirement of backwards compatible operation mode is presented, which means that the LTE-Advanced systems should support both the multi-CC capable users and the legacy users that operate with only one CC.

1.2.2 Radio Resource Management in LTE-Advanced

RRM is used in LTE-Advanced to ensure that the radio resources are efficiently utilized, taking advantage of the available adaptation techniques [21]. A short introduction is provided here.

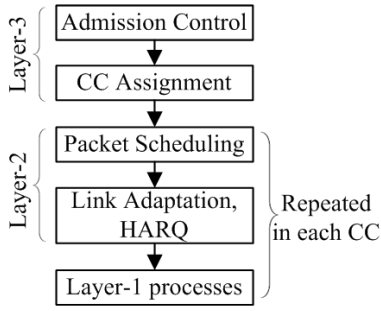


Fig. 1.2: Radio resource management in a multi-CC LTE-Advanced system.

A flowchart of RRM in LTE-Advanced is depicted in Fig. 1.2. It contains admission control and CC assignment in layer-3, packet scheduling, link adaptation and Hybrid Automatic Repeat Request (HARQ) management in layer-2, as well as

the layer-1 processes. In a system with multiple CCs, the layer-2 and layer-1 operations are repeated within each CC. Nevertheless, it is possible to have the cooperation among multiple CCs to optimize the system performance.

A user will be served by a base station only if it is admitted by the admission control. The admission decision is made based on the Quality of Service (QoS) requirements of the user, the channel quality, the current cell load, etc. After a user is admitted to a cell, it will be provided with a single or multiple CCs, depending on its requirement and the terminal type. This is achieved by layer-3 CC assignment, which is a new functionality in LTE-Advanced.

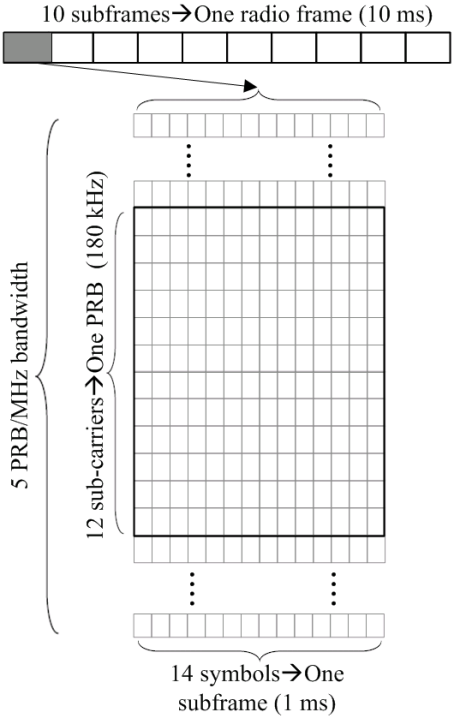


Fig. 1.3: Time-frequency domain frame structure of LTE systems.

Fig. 1.3 illustrates the time-frequency grid resource in LTE systems [22, 23], which is also equivalent to one CC of LTE-Advanced. One Physical Resource Block (PRB) is the minimum resource element that can be assigned to a user. It consists 12 consecutive sub-carriers, and has a bandwidth of 180 kHz. One PRB also corresponds to a subframe in the time domain, which contains 14 Orthogonal Frequency Division Multiplexing (OFDM) or Single-carrier Frequency Division Multiplexing (SC-FDM) symbols, with a Transmission Time Interval (TTI) of 1 ms. There are 10 subframes in a radio frame, and the total duration is 10 ms. The available resources can be shared among multiple users within each TTI, using

packet scheduling. Different system requirements, e.g., maximization of the cell throughput, cell edge user throughput, or fairness, can be satisfied by using a proper scheduling metric.

Link adaptation is part of the RRM in layer-2. By selecting the proper modulation and coding scheme, it tries to maximize the spectral efficiency while satisfying a certain Block Error Rate (BLER) constraint. HARQ management is applied to improve the performance by combining the retransmission with previous transmission(s). After assigning a user with resources, and selecting the proper modulation and coding scheme (and potentially the Multiple Input Multiple Output (MIMO) mode), independent layer-1 transmissions are carried out at each CC.

1.2.3 Feedback Report

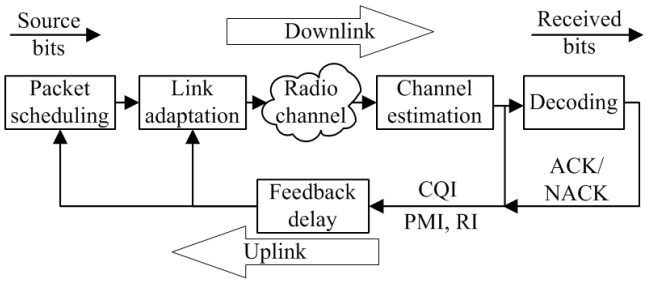


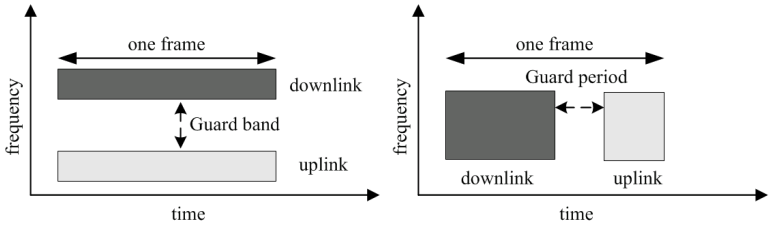
Fig. 1.4: A downlink wireless communication system with uplink feedback, channel aware packet scheduling, and link adaptation.

To benefit from the frequency domain diversity, the knowledge of channel quality at the transmitter is required [24–26]. A typical wireless communication system with feedback link is depicted in Fig. 1.4, where the receiver estimates the Channel Quality Information (CQI) and sends it back to the transmitter, together with the Acknowledgement (ACK)/Non-acknowledgement (NACK) of the packet reception [27]. When codebook based MIMO transmission is used, the base station also needs the information on which precoding matrix to use (the Precoding Matrix Indicator (PMI)), and the number of supported data streams (the Rank Indication (RI)) [27].

1.2.4 FDD and TDD Mode

The duplexing of uplink and downlink transmissions is normally carried out in frequency or time domain and is known as Frequency Division Duplexing (FDD)

or Time Division Duplexing (TDD), respectively [12]. FDD-based systems typically employ paired channels for uplink and downlink transmissions. Due to the symmetry in the two channels, FDD is suitable for voice transmission, where the uplink/downlink traffic is symmetric. Therefore, it is extensively used in the 2G and 3G networks. Nowadays the data traffic contributes to most of the total traffic volume. This data traffic is usually asymmetric in the downlink and uplink and requires different amount of resources in the two directions [28]. TDD has attracted much interest from a research point of view because it allows uplink and downlink transmission to share the same channel at different times, and thus can be adapted according to the traffic condition [29–31].



(a) Frequency division duplexing (b) Time division duplexing

Fig. 1.5: Duplexing modes to support uplink and downlink transmissions.

An FDD system with paired uplink/downlink channels is depicted in Fig. 1.5(a), whereas a TDD system is shown in Fig. 1.5(b), with uplink/downlink separated in the time domain. FDD uses separate channels for the uplink and downlink transmissions. Therefore it is immune to the uplink-downlink mutual interference [32]. TDD requires a guard period to separate the two transmissions. If the transmissions from neighboring cells are not fully aligned, a base station may receive interference from the downlink transmission of its neighboring base stations, and vice versa. In many situations, this mutual interference significantly penalizes the uplink performance due to the relatively low uplink transmit power.

1.2.5 Wide Area and Local Area Networks

The wide area network is, as the name implies, characterized by the large coverage. It is equipped with elevated outdoor base stations transmitting at high power. To increase the data rate and reduce the interference, a wide area site is usually sectorized into several (typically 3) sectors, each served by a directional antenna. The base stations have fixed positions, thereby it is possible to have a pre-planning of the network, including frequency channel assignment, transmit

power coordination etc. Most of the existing wireless networks are deployed in wide area, and they are known as the macro-cells.

In recent years, along with the various wireless multimedia applications, the amount of traffic generated from indoor usage has increased rapidly. It is found that more than 70% of the data traffic originates from within buildings [33], and the traditional wide area networks may not be able to satisfy the user requirement in areas with high user density [34]. In this case, local area networks can be deployed to serve a small number of users in geographically limited areas [35], e.g., the indoor home and office scenarios. In a local area network, the neighboring base stations are close to each other, and the inter-cell interference plays a vital role in the system performance [36]. One of the popular base station types for local area is the so-called user-deployed home base stations, which basically means uncoordinated deployment without any prior network planning. Moreover, due to the small number of users that are served per base station, the traffic condition changes dramatically in a local area network. All these aspects point to a self-organized operation of the local area networks. Fig. 1.6 shows a local area network lying in a traditional wide area network. It serves the users within a limited area, using an omni-directional antenna at low transmit power.

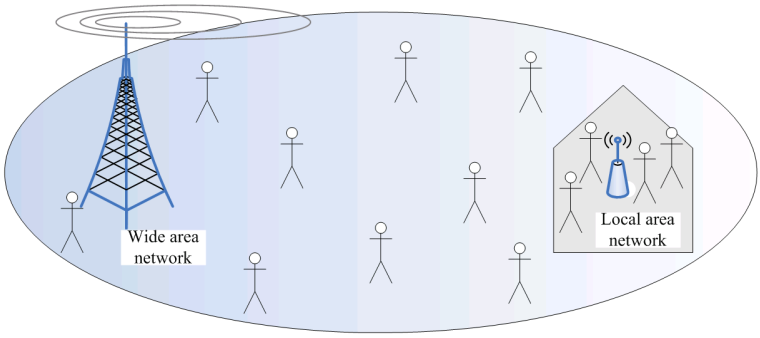


Fig. 1.6: Wide area and local area networks.

1.3 State of the Art

TDD usually requires time synchronization among neighboring cells. Many techniques have been developed to offer different synchronization levels, e.g., Institute of Electrical and Electronics Engineers (IEEE) 1588 [37], Global Positioning System (GPS) [38] and firefly based synchronization algorithms [39–42]. Similar techniques have been considered as candidates for LTE-Advanced [43]. However, the effect of time synchronization on system performance has been evaluated

mainly in the wide area networks. Since the interference condition of the local area network differs from the wide area network, a similar study for the local area network is needed to select the appropriate synchronization technique.

Frequency reuse has been the object of extensive studies for the existing systems to improve performance [44, 45]. Usually, frequency reuse 1 is preferable in wide area networks. However, in local area networks, a high reuse factor that avoids strong interference from a close interfering node might be beneficial. Therefore, it is necessary to investigate the performance of frequency reuse in a local area network. Meanwhile, the local area network usually lacks the central controller, and hence self-organized algorithms should be studied for frequency sharing [46–50].

The migration from single to multi-CC systems has been studied for High Speed Downlink Packet Access (HSDPA) [51, 52], and Code Division Multiple Access (CDMA) [53–58] systems. It is found that load balancing is essential for the system performance, and many techniques have been developed to balance the load across the CCs by selecting an appropriate CC for each user. Besides the CC selection, the resources should also be properly shared among users within each CC. In a system with both single CC and multi-CC capable users, how to improve fairness among the users along with system efficiency is a challenge.

To reduce the amount of resources occupied by the feedback report, many techniques have been developed. E.g., by reporting a quantized value instead of the full channel state [59–61], using best-M report [27, 62–64], or increasing the feedback interval [65]. The ACK/NACK signaling can be reduced by bundling the ACK/NACKs of one user [27, 66]. Its has been shown to give attractive performance in an LTE TDD system [67]. In a multi-CC system, the requirement of feedback is even more challenging than that already exist in a single-CC system. Therefore, the feedback reduction mechanism should be carefully designed in such multi-CC systems.

1.4 Problem Delineation

This study deals with the system level management of multiple CCs in LTE-Advanced. Both local area and wide area networks are taken into consideration.

The local area system has a different interference condition from the wide area, and it is expected that the performance can be improved by basic interference coordination among cells. This hypothesis will be looked into specifically with respect to network time synchronization and frequency sharing. A TDD system will be used in order to investigate the impact of time synchronization on system performance. In the other studies of the thesis, the interference between uplink and

downlink is not considered and hence the findings and proposals are valid for both FDD and perfectly synchronized TDD systems. The self-organized operations will be prioritized for frequency sharing.

For the wide area network, much lower gain is expected from cell level frequency sharing than in the local area. Therefore, focus is put to the resource allocation within each cells. The study will address the problem of load balancing across the aggregated CCs, and fairness provision among users that can access different number of CCs. The potential of reducing uplink feedback overhead will also be analyzed both within each CC and across the CCs, and the trade-off between downlink performance and uplink overhead reduction will be made.

1.5 Research Methodology

The multi-CC management involves various operations in different network layers, which are interacting with each other. Therefore it is difficult to derive the stringent mathematical optimal solution to maximize the system performance. In this study, a heuristic approach is therefore used. Some suitable algorithms are developed for each of the investigated topics, and their performance is evaluated and compared with each other. The interaction among these steps also adds difficulties in the mathematical modeling and the theoretical estimation of the system performance. Therefore, the performance is mainly evaluated relying on system level simulations. Some simple theoretical analyses are also made when possible, in order to verify the simulation results.

In order to investigate the interaction between uplink and downlink in a TDD system, transmissions in both directions are simultaneously simulated, with full transmit power in the downlink, and fractional path loss compensated uplink transmit power. The local area study is mainly to analyze the potential gain of having network synchronization and spectrum sharing, which does not require high accuracy. Therefore, several modeling simplifications have been made. E.g., a modified Shannon formula [68], which includes various system imperfections, is used for estimating the system throughput based on the Signal to Interference plus Noise Ratio (SINR) condition. A channel blind scheduler is used for the packet scheduling, and the fast fading channel is not modeled. This is so because the indoor channel varies slowly, and there is little frequency domain diversity to exploit.

Since the focus of the wide area study is on resource sharing among users within a cell, a detailed system model is needed. A simulator that is fully aligning with the LTE specifications is built, and its accuracy is higher than the one used for local area. It includes the modeling of admission control, carrier load balancing, packet scheduling, link adaptation, HARQ, channel estimation, packet reception,

periodical feedback with transmission delay, path loss, shadowing and fast fading channels, etc.

1.6 Contributions and Recommendations

The first topic of research is the performance evaluation of a local area TDD system with different levels of network time synchronization. It is suggested to achieve at least loose network synchronization. Otherwise, the uplink transmission will be severely affected by the downlink interference, giving very poor performance. The following article has been published based on this finding:

- Yuanye Wang, Simone Frattasi, Troels B. Sørensen and Preben Mogensen, "Network time-synchronization in TDD based LTE-Advanced systems," *in Proc. IEEE VTC*, Apr 2009.

Next, the benefit of frequency reuse in local area networks is analyzed, and frequency reuse-2 is found to offer several-fold performance gain in cell edge user throughput, with little or no loss in average cell throughput. Therefore, spectrum sharing among local area cells is recommended. Several distributed algorithms are developed to facilitate cell-level spectrum sharing. The following articles related to the spectrum sharing in local area networks have been published:

- Sanjay Kumar, Yuanye Wang, Nicola Marchetti, István Kovács, Klaus Pedersen and Preben Mogensen, "Spectrum load balancing for flexible spectrum usage in local area deployment scenario," *in Proc. IEEE DySPAN*, Oct 2008.
- Yuanye Wang, Sanjay Kumar, Luis Garcia, Klaus Pedersen, István. Kovács, Simone Frattasi, Nicola Marchetti and Preben Mogensen, "Fixed frequency reuse for LTE-Advanced systems in local area scenarios," *in Proc. IEEE VTC*, Apr 2009.
- Luis Garcia, Yuanye Wang, Simone Frattasi, Nicola Marchetti, Preben Mogensen and Klaus Pedersen, "Comparison of spectrum sharing techniques for IMT-A systems in local area networks," *in Proc. IEEE VTC*, Apr 2009.
- Yuanye Wang, Nicola Marchetti, István Kovács, Preben Mogensen, Klaus Pedersen and Troels Sørensen, "An interference aware dynamic spectrum sharing algorithm for local area LTE-Advanced networks," *in Proc. IEEE VTC*, Sep 2009.
- Sanjay Kumar, Yuanye Wang and Nicola Marchetti, "Self-organized spectrum chunk selection algorithm for local area LTE-Advanced," *in Proc. ITU-T Kaleidoscope 2010*, Dec 2010.

For the wide area network, balancing the load across multiple CCs is essential to improve the multi-user diversity. The best balancing method is identified among several known ones. In order to maximize the network utility and improve the fairness among users, a cross-CC packet scheduler is developed. It improves the cell edge user throughput without penalizing the average cell throughput. A simple theoretical model is also developed to verify the simulation results. Findings from this study are summarized in the following articles:

- Yuanye Wang, Klaus Pedersen, Preben Mogensen and Troels Sørensen, "Carrier load balancing methods with bursty traffic for LTE-Advanced systems," *in Proc. IEEE PIMRC*, Sep 2009.
- Yuanye Wang, Klaus Pedersen, Preben Mogensen and Troels Sørensen, "Resource allocation considerations for multi-carrier LTE-Advanced systems operating in backward compatible mode," *in Proc. IEEE PIMRC*, Sep 2009.
- Yuanye Wang, Klaus Pedersen, Troels Sørensen and Preben Mogensen, "Carrier load balancing and packet scheduling for multi-carrier systems," *in IEEE Trans. Wireless Communications*, May 2010.
- Klaus Pedersen, Luis Garcia, Hung Nguyen, Yuanye Wang, Frank Frederiksen and Claudio Rosa, "Carrier aggregation for LTE-Advanced: functionality and performance aspects," *accepted for IEEE Communications Magazine*.
- Yuanye Wang, Klaus Pedersen, Troels Sørensen and Preben Mogensen, "Utility maximization in LTE-Advanced systems with carrier aggregation," *accepted for IEEE VTC*, May 2011.
- Hung Nguyen, István Kovács, Yuanye Wang and Klaus Pedersen, "Downlink performance of a multi-carrier MIMO system in a bursty traffic cellular network," *accepted for IEEE ICC*, Jun 2011.

On the feedback reduction, several proposals have been made to reduce the CQI and ACK/NACK overhead both within a CC and across multiple CCs. These proposals can be found in:

- Klaus Pedersen and Yuanye Wang, "Downlink packet scheduling and uplink feedback configuration for LTE-Advanced," *submitted patent application*, Jun 2009.
- Yuanye Wang, Klaus Pedersen, Miguel Navarro, Preben Mogensen and Troels Sørensen, "Uplink overhead analysis and outage protection for multi-carrier LTE-Advanced systems," *in Proc. IEEE PIMRC*, Sep 2009.
- Klaus Pedersen and Yuanye Wang, "Assignment of component carriers," *submitted patent application*, Feb 2010.

- Yuanye Wang, Klaus Pedersen, Troels Sørensen and Preben Mogensen, "Downlink transmission in multi-carrier systems with reduced feedback," *in Proc. IEEE VTC*, May 2010.
- Yuanye Wang, Klaus Pedersen, Troels Sørensen and Preben Mogensen, "System level analysis of ACK/NACK bundling for multi-component carrier LTE-Advanced," *accepted for IEEE VTC*, May 2011.

Furthermore, a load-adaptive average best-M method is developed for the CQI feedback, which achieves better performance than fixed best-M. It is also observed that the behavior of the Generalized Proportional Fair (GPF) scheduler is sensitive to the traffic model. A cell-level CC assigning method is proposed to achieve different frequency reuse patterns. The user velocity aware CC assignment is also investigated. These findings are presented in the following materials:

- Klaus Pedersen and Yuanye Wang, "Terminal speed dependent component carrier load balancing," *submitted patent application*, Oct 2009.
- Klaus Pedersen and Yuanye Wang, "Fractional frequency reuse in cellular communication systems with multiple component carriers," *submitted patent application*, Nov 2009.
- Pablo Ameigeiras, Yuanye Wang, et al., "Traffic models impact on OFDMA systems design," *submitted to the Springer journal of Mobile Communication, Computation and Information.*

Besides the main emphasis on system level analysis of multi-CC management in LTE-Advanced, many other related topics have been studied, including MIMO precoding, link adaptation, etc. These studies have been published in:

- Troels Sørensen, Daniel Figueiredo, Yuanye Wang, et al., "Description of the network optimal transmit and receive," *IST-SURFACE D5*, Jan 2008.
- Troels Sørensen, Yuanye Wang, et al., "First results from link and system level simulations," *IST-SURFACE D7.3*, Jan 2008.
- Yuanye Wang, et al., "Allocation fairness for MIMO precoded UTRA-LTE TDD system," *in Proc. IEEE VTC*, Sep 2008.
- Muhammad Rahman, Yuanye Wang, et al., "Channelization issues with fairness considerations for MU-MIMO precoding based UTRA-LTE/TDD systems," *in Proc. IEEE VTC*, Sep 2008.
- Suvra Das, Muhammad Rahman and Yuanye Wang, "Hybrid strategies for link adaptation exploiting several degrees of freedom in WiMAX systems," *book chapter in "WiMAX Evolution: Emerging Technologies and Applications"*, Wiley, Jan 2009.

As a short summary, during this PhD study, one IEEE transaction paper was published; one magazine paper has been accepted for IEEE Communication Magazine; one journal paper has been submitted; one book chapter has been published by Wiley; 12 conference papers have been published, and 3 are accepted for publication; 4 patent applications have been submitted by the Nokia Siemens Networks patent office. The findings also contribute to two European Union project deliverables and many Nokia Siemens Networks internal deliverables.

Other than the publications, one simulator for local area LTE-Advanced system was built, which has been used in many master student projects. It is also used by some of the colleagues, contributing to 3GPP meetings and many other projects. For investigating the performance of wide area systems, a multi-CC system simulator was developed. It is based on a single-CC LTE-Rel'8 simulator from Nokia Siemens Networks. The extension from single to multiple CCs requires many new functionalities, including different carrier aggregation patterns, CC assignment, inter-CC CQI reduction, ACK/NACK bundling. Meanwhile, modifications are needed throughout the existing simulator, e.g., the packet scheduler, fading channel, traffic model, admission control, channel estimation, updating/collecting statistics. The multi-CC LTE-Advanced simulator is also used by other colleagues for the Nokia Siemens Networks activities in 3GPP and other research projects.

1.7 Thesis Outline

The thesis is divided into 6 chapters and 5 appendices, where the first chapter is the introduction and the last chapter concludes the work. Chapters 2 and 3 are devoted to local area network, which aim to provide the benchmarking results and the general trends of different techniques in this scenario. Chapters 4 and 5 focus on the downlink transmission of a wide area network with feedback in the uplink. A brief description of the chapters is provided below:

- Chapter 2: Network time synchronization in local area TDD systems — The effect of network time synchronization, and hence different levels of uplink/downlink alignment is investigated.
- Chapter 3: Flexible spectrum usage in local area LTE-Advanced networks — This chapter first shows that the local area network performance can be significantly improved by frequency reuse, then develops several distributed algorithms to achieve similar performance as frequency reuse.
- Chapter 4: Carrier load balancing and packet scheduling for wide area multi-CC Systems — The performance of resource allocation in wide area networks with multi-CCs is addressed in this chapter. Several load balancing methods are studied, and the best one is identified. A cross-CC packet scheduler is

proposed to improve the system performance as compared to the case of using independent scheduling per CC.

- Chapter 5: Downlink performance of multi-CC systems with reduced feedback — Several possibilities to reduce the feedback overhead while maintaining good downlink performance are studied. Selecting the proper reduction technique based on the cell load is suggested.

- Chapter 6: Conclusions and future work — A summary of the overall study is provided. Based on the findings, the system level recommendations for the deployment of LTE-Advanced in both local area and wide area are given. Some possible future works are also listed.

In order to support the work, five appendix chapters are provided:

- Appendix A: Scenario dependent simulation assumptions — The Simulation assumptions for both local area and wide area are summarized in this appendix. The modeling of antenna sectorization, wireless channel, link adaptation, HARQ management, receiver combining and traffic model is also presented.

- Appendix B: Capacity of MIMO systems and the multi-user gain — The capacity of a MIMO system using channel blind or channel aware packet scheduling is derived mathematically. The multi-user gain is observed to follow a logarithmic function versus the number of users per cell. This logarithmic approximation of the multi-user gain is used for the performance estimation in other chapters.

- Appendix C: Theoretical analysis of the finite buffer transmission — A theoretical model of the finite buffer traffic is presented, based on which the performance estimation can be made. The estimation and analysis confirms the implementation of the finite buffer traffic, the bandwidth extension using carrier aggregation and the carrier load balancing is correct, and also verifies the accuracy of the simulation results in Chapter 4 and Chapter 5.

- Appendix D: Further improvement of the wide area network performance — Several techniques to further improve the system performance are analyzed. Their benefits and limitations are also presented.

- Appendix E: Supplementary results for previous chapters — This chapter provides additional results to Chapter 2, 3 and 5 on the topics of network time synchronization with Dynamic TDD (D-TDD), self-organized frequency reuse and CQI reduction using the full buffer traffic model.

Chapter 2

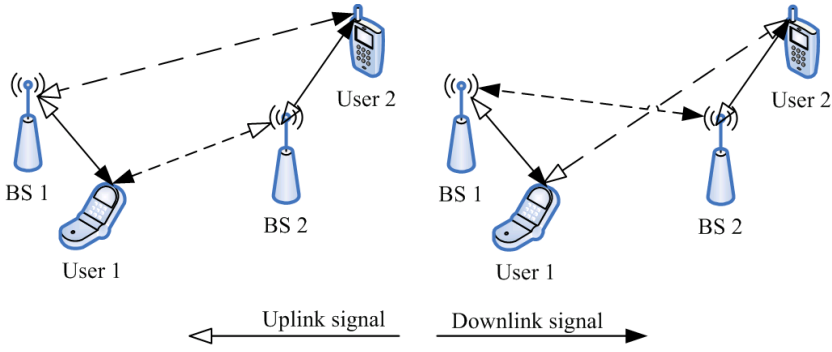
Network Time Synchronization in Local Area TDD Systems

2.1 Introduction

In recent years, the amount of traffic generated from indoor usage has increased rapidly. It is found that more than 70% of the data traffic originates from within buildings [33], and the traditional wide area networks may not be able to satisfy the user requirement in user dense areas [34]. The local area networks can be deployed to serve a small number of users in geographically limited areas [35], e.g. the indoor home and office scenarios. The data traffic in local area networks is usually asymmetric in the downlink and uplink [28], therefore the duplexing scheme of Time Division Duplexing (TDD) is attractive [29–31].

Although TDD can adapt to the traffic condition, it has its own disadvantages, the most critical one is the mutual interference between uplink and downlink transmission. Fig. 2.1 depicts the cases with isolated or overlapped uplink/downlink transmissions in a two-cell scenario with one user per cell. In this figure, the solid lines are used to indicate the intended signals; the dashed lines are for interference; 'BS' is short for 'base station'. If the uplink and downlink are isolated, a user will receive interference from its neighboring base stations in the downlink; in the uplink, interference is generated by the users in neighboring cells. This is shown in Fig. 2.1(a). When the uplink and downlink transmissions overlap, e.g., one cell is in downlink mode whereas the other one is in uplink mode, the interference source

will change, as can be seen from Fig. 2.1(b). Note that, if the uplink/downlink alignment is imperfect, transmissions in the two directions are sometimes isolated across cells, but overlapped in another time instant. Therefore, the interference pattern will switch between the two cases shown in Fig. 2.1.



(a) Isolated transmission: both in uplink or downlink mode.

(b) Overlapped transmission: one in uplink, the other in downlink mode.

Fig. 2.1: Impact of uplink/downlink alignment on the inter-cell interference.

The alignment between uplink and downlink can be improved by achieving network time synchronization. The effect of network time synchronization on the system performance has been extensively studied in the wide area, showing severely degraded uplink performance when synchronization is lost [69, 70]. It is also claimed in [71–74] that achieving network synchronization is needed in the wide area network. Many techniques have been developed for the purpose of achieving time synchronization. E.g., IEEE 1588 standard defines a protocol for precise clock synchronization in a network [37]. It works based on a master-slave relationship in which each slave synchronizes to its master. This standard has been widely used in both wired networks [75] and wireless networks, e.g., Wireless Local Area Network (WLAN) [76]. The Global Positioning System (GPS) can also be used to provide the common reference clock within a network [38]. Firefly based synchronization algorithms [39–42] operate in a distributed manner and can be useful for cases where a master clock is not available. IEEE 1588v2, GPS and network listening techniques have been considered as candidates for LTE-Advanced [43].

In this chapter, focus is put on the performance comparison of different synchronization levels in the local area: full synchronization, loose synchronization and un-synchronization. The meanings of these level will be introduced in the next section. The purpose is to estimate the gain obtained by achieving network synchronization, so as to justify the use of these techniques in the local area network.

The rest of the chapter is organized as follows: Section 2.2 describes the synchronization levels and their impacts on the network transmission pattern. Section 2.3 presents the simulation results in small home and office scenarios. In Section 2.4, the performance in large home and office scenarios with various activity factors is shown. Finally, Section 2.5 concludes the whole chapter.

2.2 Duplexing and Network Synchronization Types under Consideration

In this study, three levels of network synchronization are considered, including:

Full synchronization: all base stations have the same reference clock. It requires the accuracy of a few microseconds and can be obtained by using IEEE 1588 protocol or GPS.

Loose synchronization: some residual error in the alignment of the clocks between base stations is allowed, up to a few milliseconds. It can be achieved in a distributed manner. For this study, this residual error is introduced in terms of the maximum mismatch in the start time of a radio frame relative to a common reference clock. According to the LTE specification [23], one radio frame contains 10 subframes, with a duration of 1 ms each. These subframes can be configured for uplink or downlink transmission, depending on the traffic requirement. In this study, a reference clock residue error of up to 1 ms is assumed, which corresponds to either 0 or 1 subframe shift.

Un-synchronization: each base station has its own clock, independent of the other base stations. Therefore, uplink and downlink transmissions have the highest possibility of mutual interference as compared to the other two synchronization levels.

A TDD system can be configured with different radio frame patterns to support different ratios of the uplink and downlink traffic. According to [23], there are 7 possible patterns in LTE, providing uplink to downlink ratios from 1:9 to 3:2. Depending on the uplink/downlink ratios of the cells within the network, the TDD systems can also be classified into two different modes [77]:

Static TDD (S-TDD): the same uplink-to-downlink ratio is used in all cells.

D-TDD: the portions of uplink and downlink transmissions are assigned dynamically in different cells.

The D-TDD system offers better adaptation to the uplink/downlink traffic than S-TDD. However, due to the dynamic partitioning of the subframes, it is impossible to fully isolate the uplink and downlink transmissions. Even with full synchronization, D-TDD may still suffer from downlink interference in the uplink transmission, and vice versa [77]. S-TDD, on the other hand, offers better protection against the uplink/downlink mutual interference but has the drawback that the sharing of uplink and downlink transmissions may not match the traffic requirement of each cell.

In this chapter, all cells are assumed to have the same uplink to downlink ratio of 50-50%, which is similar to the uplink-downlink configuration 6 in [23] and corresponds to the S-TDD mode. Fig. 2.2 shows the transmission patterns with different synchronization levels, where the dark subframes are used for downlink transmission, and the grey ones are for uplink transmission. The performance of D-TDD with different synchronization levels is summarized in Appendix E.1.

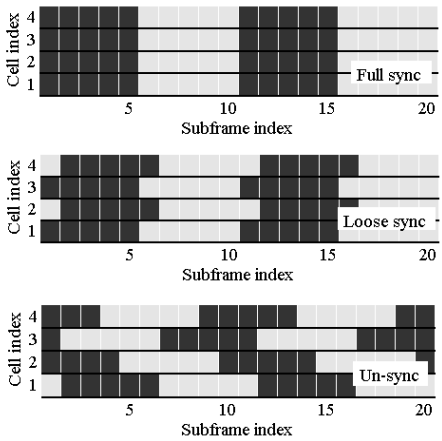


Fig. 2.2: Transmission patterns with S-TDD and different synchronization levels.

2.3 Simulation Results in Small Indoor Scenarios

This section and the following one provide simulation results for different synchronization levels in the indoor home and office scenarios. The simulation methodology and assumptions are in general according to the LTE-Advanced setup, and are detailed in Appendix A.1.

The effect of network synchronization highly depends on the relative interference level coming from the downlink and uplink directions, which is ultimately de-

termined by the uplink and downlink transmit power. While full power of 24 dBm is used in the downlink, the uplink power is subject to fractional path loss compensation according to (A.4). Fig. 2.3 provides the Cumulative Distribution Function (CDF) of the path loss dependent uplink transmit power in both home and office scenarios. As shown, the average uplink transmit power is as low as -13 dBm in the home scenario and -4 dBm in the office scenario, which is much lower than in the downlink. Note that the uplink transmit power level is the same in both small and large scenarios, as it depends only on the path loss value to the serving base station.

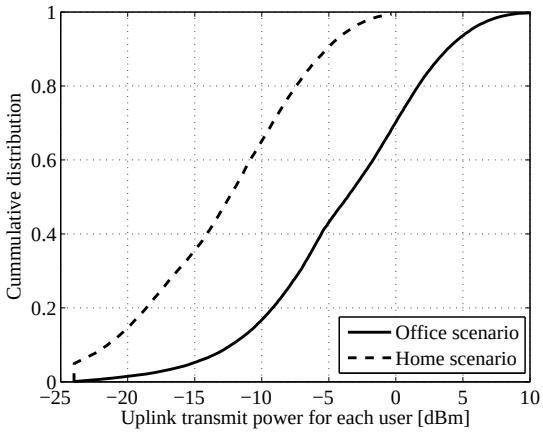


Fig. 2.3: CDF of uplink transmit power in home and office scenarios.

The downlink performance in small indoor scenarios is summarized in Fig. 2.4. From this figure it can be seen that, in both scenarios, when synchronization is achieved, the performance is actually worse than without synchronization (∼30% loss in average cell throughput and 50∼75% loss in cell edge user throughput). The reason is that power control is used in the uplink, and hence the average interference generated in the uplink is lower than that generated in the downlink. Un-synchronization benefits from the very low transmit power in the uplink direction.

Although not beneficial in the downlink, achieving synchronization brings a huge gain in the uplink, as shown in Fig. 2.5. In the home scenario without synchronization, 50% of the users cannot get any service, while a perfectly synchronized system at the same CDF point (50%) can already achieve a user throughput of around 25 Mbps. Similar behavior is also observed in the office scenario: 30% of the users cannot be served when synchronization is lost, whereas full synchronization offers 7 Mbps user throughput (30-percentile worst). The different impacts of time synchronization on the uplink and downlink transmissions are again due to

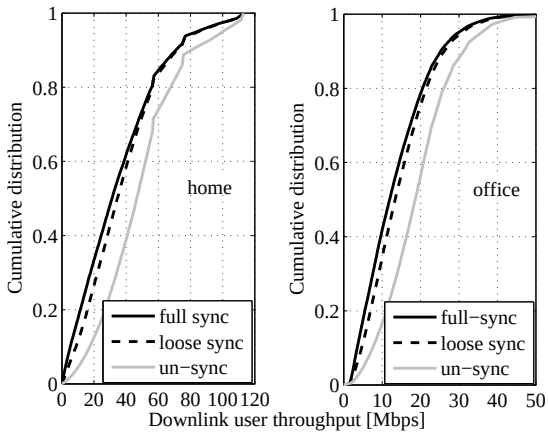


Fig. 2.4: CDF of downlink user throughput with different synchronization levels in small indoor scenarios.

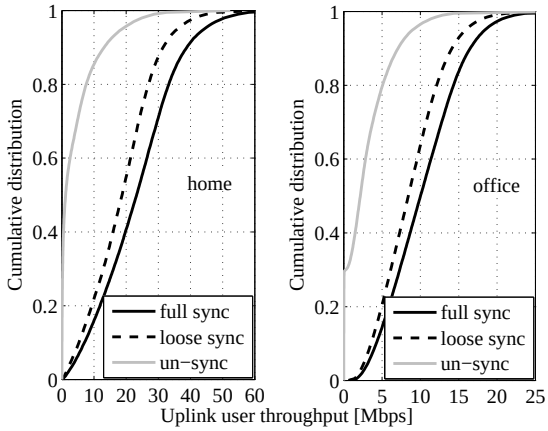


Fig. 2.5: CDF of uplink user throughput with different synchronization levels in small indoor scenarios.

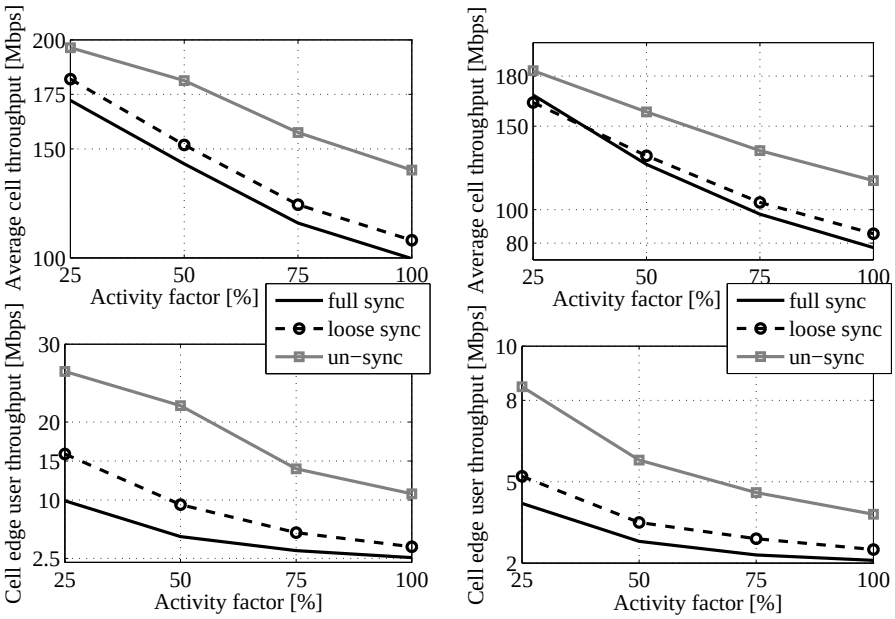
the considerably lower transmit power in the uplink direction than in the downlink (see Fig. 2.3). It is also observed that the home scenario is more affected by network time synchronization than the office scenario, because the former has lower uplink transmit power, and hence higher difference between uplink and downlink interference level.

It is worth noting that ideal control channel is assumed in this study. In reality, however, the uplink control channel also suffers from the loss of network time syn-

chronization. It carries the channel information and reception acknowledgement for the downlink. When the uplink transmission is heavily interfered by the downlink signals, the downlink performance will also be penalized due to the lack of reliable uplink feedback. In this situation, achieving network time synchronization will increase the reliability of the feedback compared to un-synchronization, and consequently improve the downlink performance.

Loose synchronization has a better performance than full synchronization in the downlink direction. In terms of uplink performance, it achieves around 80% of the performance with full synchronization. This is due to the fact that the subframe suffering from downlink interference has very low SINR and can hardly be used to convey any data traffic.

2.4 Simulation Results in Large Indoor Scenarios

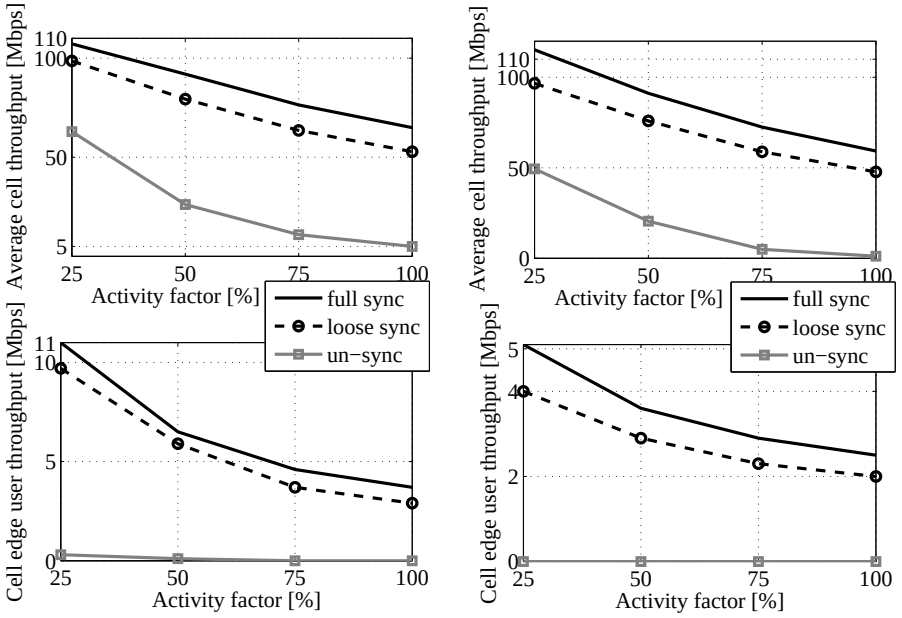


(a) Home scenario

(b) Office scenario

Fig. 2.6: Downlink performance versus activity factor. Different synchronization levels are evaluated in large indoor scenarios.

When the network contains more base stations, the loss of network time synchronization is expected to cause even worse uplink performance. The large home and office scenarios are used for this evaluation, with 16 cells each. Furthermore, the active/idle base stations are also considered, which is controlled by the activity factor. E.g., an activity factor of 50% means that on average half of the cells are active, while the other half are idle and are invisible to the active nodes.



(a) Home scenario

(b) Office scenario

Fig. 2.7: Uplink performance versus activity factor. Different synchronization levels are evaluated in large indoor scenarios.

Fig. 2.6 and Fig. 2.7 show the average cell throughput (the upper subplot) and cell edge user throughput (the lower subplot) in the large indoor scenarios for the downlink and uplink transmission, respectively. Similar as in the small indoor scenarios, achieving network synchronization benefits the uplink transmission (Fig. 2.7), but not the downlink transmission (Fig. 2.6). Looking at the uplink performance in Fig. 2.7, loose synchronization achieve ∼80% of the performance of full synchronization in both average cell and cell edge user throughput. Without network synchronization, very poor average cell throughput is experienced, and cell-edge (5-percentile worst) users could hardly get any service.

To have a clear view of the performance without synchronization, the CDF of the

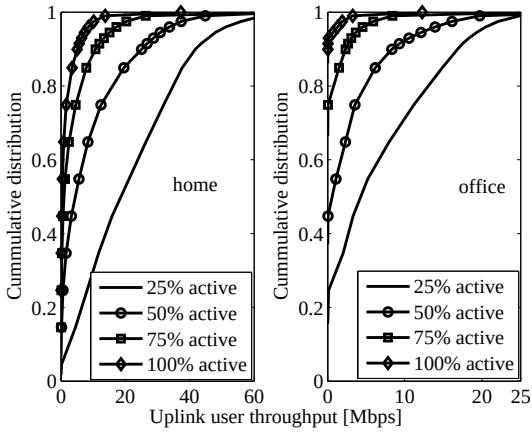


Fig. 2.8: CDF of uplink user throughput without network time synchronization in large indoor scenarios.

uplink user throughput is plotted in Fig. 2.8 for the un-synchronized case. From this figure it can be see that, as the activity factor increases, more base stations will be in active mode and interfere with each other. As a consequence, the uplink performance significantly decreases with the activity factor, and more users cannot be served. It is also observed that the office scenario is more penalized by the loss of network synchronization. This is in contradiction to the result obtained in small indoor scenarios. The reason is that, in the large office scenario, each base station is exposed to three Line of Sight (LOS) base stations. These base stations can potentially generate high interference when synchronization is lost. On the other hand, there is only one LOS interfering base station in the small office scenario and zero in the home scenario, therefore the performance in these scenarios is less impacted by the loss of network time synchronization than the large office scenario.

2.5 Conclusion

In this chapter the effect of network time synchronization in local area TDD networks has been investigated. The evaluation is done using system level simulations based on the LTE-Advanced specifications. Both indoor home and office scenarios are used for the evaluation.

According to the obtained results, it is found that the loss of network time synchronization leads to unacceptable poor uplink performance. However, the downlink transmission actually gains from avoiding interference in the same direction (when

synchronization is not perfect). This is due to the fact that the transmit power in the uplink is on average lower than in the downlink. Despite the loss in downlink transmission, it is suggested to achieve at least loose network time synchronization to guarantee reasonable uplink service. It is worth noting that the downlink power control has also been considered for local area networks [22]. If it is applied, and the average transmit power is similar in the two directions, both uplink and downlink will be affected by the loss of time synchronization. In most cases, the SINR covers a wider range when synchronization is not achieved, and the cell edge user throughput will be penalized.

Chapter 3

Flexible Spectrum Usage in Local Area LTE-Advanced Networks

3.1 Introduction

The spectrum is a scarce resource that should be efficiently used to achieve high performance. In traditional wide area networks, e.g. the micro- and macro-cells, universal frequency reuse (reuse-1) is usually applied because it offers the best performance [44]. However, in the local area network, due to the small cell size and the uncoordinated placement of the Access Point (AP)s, users near the cell boundary of neighboring cells may suffer from strong interference while the received signal from the serving cell is weak [36]. Also, the interference condition is not so stable in the local area network, because each AP serves only a few users and the AP can be temporarily idled if none of the users require data transmission. Last but not least, in a local area network there is usually lack of a central controller, which is aware of the interference condition among neighboring cells and can come up with a proper frequency allocation to each cell. Considering these issues, the local area cells typically operate in the self-organized manner. In this study, the assumption of self-organization will be made, and effort is devoted to the development of algorithms for self-organization, which facilitates the spectrum sharing in local area networks.

Frequency reuse is first applied as a reference to investigate the potential gain of spectrum sharing in local area networks. Afterwards, the constraint of self-organization is followed, and hence the Flexible Spectrum Usage (FSU) technique becomes very suitable for the spectrum sharing among neighboring cells [46–50]. Several FSU algorithms have been developed to reduce the inter-cell interference, while reserving sufficient spectrum for each cell. Their performance is evaluated in various scenarios, and the pros and cons for them are presented.

The rest of the chapter is organized as follows: Section 3.2 presents the analysis and results for frequency reuse, which motivates for further study on spectrum sharing and provides the reference performance. In Section 3.3, two distributed algorithms are proposed, and their performance is summarized in Section 3.4. Conclusions of the whole chapter are drawn in Section 3.5.

3.2 Frequency Reuse: the Benchmarking Performance

In order to estimate the potential benefit of spectrum sharing in local area networks, the performance of frequency reuse is evaluated at first. For this evaluation, the frequency allocation pattern is pre-planned to minimize the inter-cell interference. The frequency plans for indoor scenarios with reuse factors of 1, 2, and 4 can be seen in Fig. 3.1[1]. By using different frequency channels in nearby cells, the strongest inter-cell interference is avoided. It significantly improves the SINR condition at the cost of reduced available spectrum per cell.

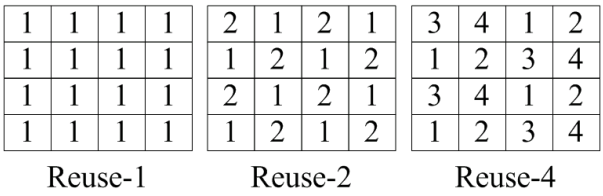


Fig. 3.1: Frequency plan for indoor scenarios with reuse factor 1, 2 and 4.

[1]For the manhattan scenario, several methods have been tested in this study for the frequency planning. The first one assigns the channel sequentially to each base station, based on the geographical location. The second one relies on the genetic algorithm [78] to compute the frequency plan. The last one is the FiDCA algorithm, which will be introduced in the next section. Based on simulation results, FiDCA offers good performance at reasonable computational complexity, and is therefore used for the frequency planning in this scenario.

3.2.1 Simple Analysis on Frequency Reuse

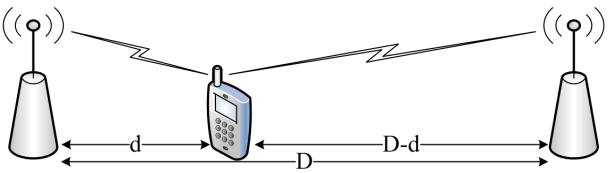


Fig. 3.2: A simple network with two APs and one user.

According to (A.3), the cell capacity with frequency reuse can be estimated as:

$$C_f = W/f * min\{BW_{eff} log_2(1 + SINR_f/SINR_{eff}), 5.02\} \tag{3.1}$$

where $f \geq 1$ is the reuse factor, which means only $1/f^{th}$ of the spectrum can be used by one cell; W is the total system bandwidth in Hz; $SINR_f$ is the achieved SINR with reuse f.

For frequency reuse to be beneficial, the increase in SINR must be able to overcome the loss in spectrum. Since the effect of SINR is scaled by the logarithm function, such a benefit can only be obtained when the SINR is not too high. In a wireless network, the SINR condition is affected by the transmit power as well as the cell size. To investigate the effect of cell size (distance between the two APs, D) on frequency reuse, a simple network with two APs and one user is considered. As shown in Fig. 3.2, the user is placed along the line between the two APs, and is connected to the AP close to it. The other AP is used to generate interference. Assuming there are 4 light walls between the user and the APs, the path loss can be calculated using the Non-Line of Sight (NLOS) model in Table A.1. Afterwards, the SINR with different reuse factors is obtained, and mapped to throughput using (3.1).

Fig. 3.3 shows the cell throughput versus the distance between APs with reuse factor of 1 or 2. Two user positions have been tested, one has a distance of $d = D/8$ to the serving base station, and the other has a distance of $D/3$. As shown in the figure, if the user is close to the serving base station ($d = D/8$), the SINR is very high so frequency reuse-1 achieves the best performance. When the user moves away from the serving base station ($d = D/3$), the SINR condition gets worse and reuse-2 may be beneficial. The cell size also affects the relative behavior of frequency reuse. In the case of $d = D/3$, reuse-2 offers better performance than reuse-1 when cell size is small ($D \leq 180$ m). Otherwise, the inter-cell interference is very weak and reuse-1 is more preferable. To conclude, the performance of frequency reuse depends on the deployment scenario. In the rest of this section, the optimal reuse factor in different scenarios will be obtained.

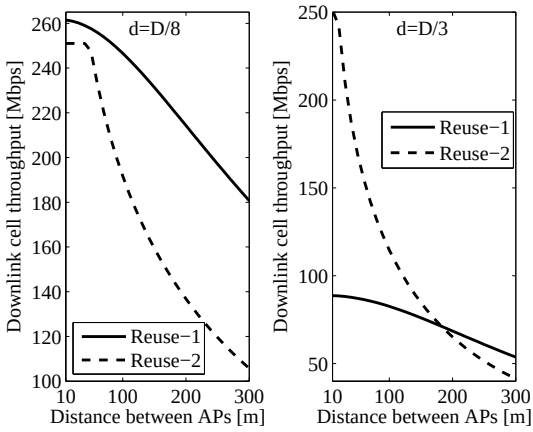


Fig. 3.3: Frequency reuse with different user positions and cell sizes.

3.2.2 Performance of Frequency Reuse in Indoor and Manhattan Scenarios

The performance of the downlink transmission in a small home scenario with 100% activity factor (all APs are activated) is shown in Fig. 3.4. From this figure it is observed that frequency reuse with factor 1 or 2 has similar average cell throughput, and is much higher than reuse-4. In terms of cell edge user throughput, reuse-2 outperforms the other reuse factors, and it achieves 6.7 times the performance of reuse-1.

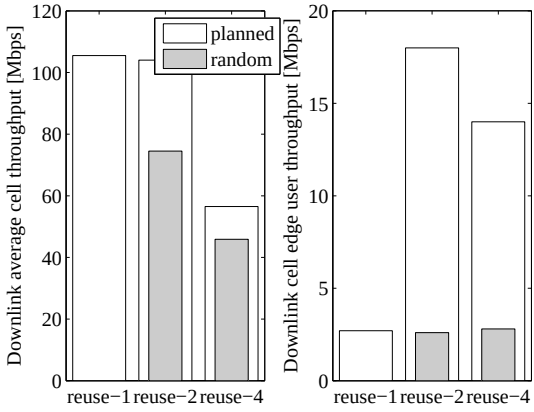


Fig. 3.4: Downlink performance with frequency reuse in a small home scenario.

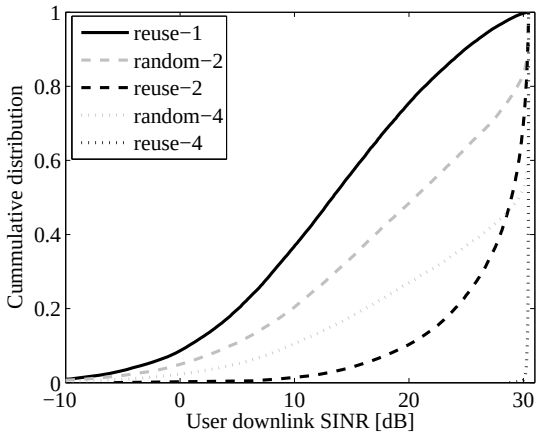


Fig. 3.5: CDF of downlink SINR with frequency reuse in a small home scenario.

The superior performance of reuse-2 can be understood from the SINR distributions for different reuse factors, as shown in Fig. 3.5. Because of the random position of the APs and the physical location based cell selection (Closed Subscriber Group (CSG)), a user may suffer from extremely high interference from its neighbors, while the signal strength is very low. This leads to the poor SINR distribution with reuse-1. With reuse-2, the 5% SINR is increased from -3 dB (reuse-1) to 16 dB. Going from reuse-2 to reuse-4, however, the gain in SINR is not so significant, because the interference power is already relatively low as compared to the signal. This small SINR gain cannot compensate for the 50% spectrum loss. Therefore, the best reuse factor in this scenario is 2.

In both Fig. 3.4 and Fig. 3.5, the performance of random channel selection is also presented, i.e., each cell randomly selects one out of the f channels. With random channel selection, the SINR improvement is not so significant as compared to planned frequency reuse. It cannot compensate for the spectrum loss and hence gives much poorer throughput than reuse-1 and planned frequency reuse.

Fig. 3.6 and Fig. 3.7 summarize the performance in all the investigated scenarios in the downlink and uplink directions. In these two figure, 'scenario-N' indicates the scenario type and the number of simulated cells. It is observed that in the indoor scenarios (home and office), frequency reuse-2 achieves the best cell edge user throughput without losing average cell throughput. Comparing the performance between the home and office scenarios, the gain of reuse-2 over reuse-1 is more significant in the home scenario than in the office scenario. This is due to the fact that in the office scenario the APs are placed at the center of each cell, and users are connected to the AP with the best signal strength (Open Subscriber Group (OSG)). Therefore the SINR condition, especially for the cell edge users, is much better than in home scenarios with randomized AP positions and CSG.

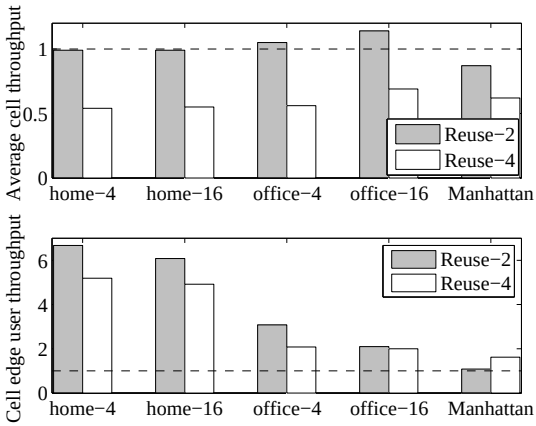


Fig. 3.6: Normalized downlink performance with respect to reuse-1.

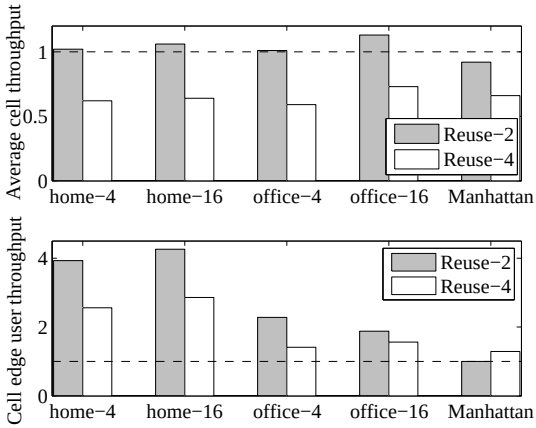


Fig. 3.7: Normalized uplink performance with respect to reuse-1.

For the outdoor Manhattan scenario, the downlink SINR distribution with different reuse factors is presented in Fig. 3.8. As can be seen from this figure, the 5% worst SINR value with reuse-1 is 0.7 dB, which is much higher than the corresponding value shown in Fig. 3.5. Moreover, the SINR improvement when using a high reuse factor of 2 is much lower than in the local area networks, because each cell suffers strong interference from many surrounding cells, and hence excluding one of these interferences gives small improvement in SINR. As a result, reuse-2 brings marginal gain in cell edge user throughput at the cost of loss in average cell throughput, and it is not recommended for the wide area Manhattan scenario. It is worth noting that, for the Manhattan scenario, the SINR improvement from

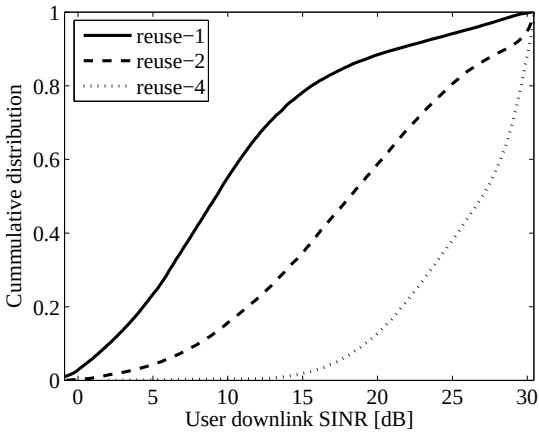


Fig. 3.8: CDF of downlink SINR with frequency reuse in the Manhattan scenario.

reuse-2 to reuse-4 is more significant than from reuse-1 to reuse-2. Consequently, the cell edge user throughput is maximized at a reuse factor of 4. However, the average cell throughput is much lower than with a low reuse factor.

3.3 FSU Algorithms to Improve the Spectrum Utilization Efficiency

The benchmarking results of frequency reuse reveal that the performance in local area networks can be significantly improved by reducing inter-cell interference. Therefore, it is worth the effort to investigate the trade-off between interference reduction and spectrum utilization. This section aims to develop self-organized FSU algorithms to approach or even surpass the performance of frequency reuse-2. These algorithms can be used without the need of central control or a priori network planning. Some of the developed algorithms during the study can be found in [79–81]. Only two of them that have better performance than the others are presented in this section.

3.3.1 Firefly Inspired Dynamic Channel Allocation

The first algorithm presented is a Dynamic Channel Allocation (DCA) algorithm that tries to achieve the optimal frequency reuse-2 pattern without the need for network planning. The term ‘channel allocation’ refers to the assignment of fre-

quency channels to each cell in a network, trying to improve the overall signal quality as well as to meet the traffic requirement of each cell.

In general there are two kinds of channel allocation algorithms: Fixed Channel Allocation (FCA) and DCA [82]. With FCA the channel allocation is fixed for each cell, which is preferable if the interference and traffic condition in a network is time insensitive. A central controller calculates the allocation pattern using a FCA algorithm and informs each cell about its channel allocation. The frequency reuse technique falls into the category of FCA. However, FCA is inappropriate for the local area networks due to the previously mentioned reasons of unstable interference condition, the dynamic traffic requirement and the lack of a central controller. The DCA algorithms allow the cells to dynamically select among available channels, according to the interference condition and the traffic requirement [83, 84]. Moreover, they can also operate in the distributed mode [82, 85, 86]. Therefore, it is advantageous to use the DCA algorithms in the local area networks. DCA has been extensively studied in the literature [83–89]. In fact, the newly developed concept of Autonomous Component Carrier Selection (ACCS) [90, 91] can also be considered as a DCA algorithm.

As mentioned before, the developed DCA algorithm in this section has the purpose of achieving frequency reuse-2 pattern in a distributed manner. It does not address the adaptation according to the traffic requirement. Instead, each cell is restricted to use half of the spectrum. The advantage of this is simplicity, and the cost is inefficient spectrum usage with very sparse number of active cells (most of the cells are idle). An extension of this algorithm to support higher reuse factors is presented in Appendix E.2. To distinguish from most of the existing work, the proposed algorithm is biologically stimulated by the firefly phenomenon. It is therefore named Firefly inspired Dynamic Channel Allocation (FiDCA).

Firefly based algorithms have been widely used to achieve network time synchronization [39–42]. They are usually based on the model of a pulse oscillator that flashes periodically and interacts with other nodes in a network. As an example, [92] developed an algorithm that when one cell 'fires', the other cells in the network adjust their clock by jumping a small step towards the mid-point of their neighbors' clock. With this algorithm, each cell keeps a record of the clock information of its neighbors, and listens to the 'firing' of the other cells. 'Fire' refers to the broadcasting of one cell about its clock, and 'jump' refers to the adjustment of the clock towards the destination.

FiDCA works in a fully distributed manner using the SINR measurement at each cell, and therefore the coordination or information exchange among the cells is unnecessary[2]. It allows a cell to adjust the channel allocation at each update

[2]To get the most benefit of FiDCA, the downlink SINR should be measured and reported to the AP for deciding the channel allocation in the downlink. A different channel might be used in the uplink, depending on the SINR measured at the AP. In this chapter, only the downlink

slot, with a pre-defined time interval between neighboring slots. Depending on the requirement of the convergence time, the update interval can vary from tens to hundreds of milliseconds, or even longer. The operation flow of FiDCA is described in the following.

Step 1: Divide the total frequency spectrum into two equal size channels. Assign a random initial status s_i to each cell, where i is the cell index, and s_i is uniformly distributed between 0 and 2. The assigned channel for cell i is $\lceil s_i \rceil$, where $\lceil \cdot \rceil$ is the ceil operation which rounds a number to the nearest integer towards infinity. The value of s_i indicates the satisfaction of one cell about its current channel allocation. If this value is close to 1, it means the cell is likely to switch to the other channel. If this value is close to the boundary, i.e., 0 or 2, it means the cell is satisfied with the current allocation and is reluctant to switch.

Step 2: At each update slot, one cell will compare the average SINR level between the current (γ_c) and the other channel (γ_o). Based on the difference (in dB), it will adjust s_i according to the following rules:

a) $\gamma_c > \gamma_o$: favors current channel, jump away from the value 1.

b) $\gamma_c \leq \gamma_o$: favors the other channel, jump towards the value 1.

The value of s_i after the jump is:

$$s_i' = (1 - \alpha)s_i + \alpha(s_i + s)/2 \tag{3.2}$$

In (3.2), $s = 0$ if the jumping favors the first channel, otherwise $s = 2$; $(s_i + s)/2$ is the mid-point between the current status and the desired operating point, which is similar to the one in [92]; α controls the speed of the jumping, it is based on the SINR gap between the two channels, a pre-defined maximum jumping step β, and the maximum SINR gap G:

$$\alpha = \beta min\{|\gamma_c - \gamma_o|/G, 1\} \tag{3.3}$$

In this step, if a cell is satisfied with the current channel allocation, it will jump towards the boundary and will be unlikely to change to the other channel; for another cell that is unsatisfied, it jumps towards the other channel, and the speed is proportional to the SINR gap between the two channels (but upper bounded by the maximum speed of β). As a result, the whole network evolves fast in the beginning, and slows down as time goes by. It stabilizes when all the cells are satisfied with their channel allocation.

Step 3: The allocated channel for cell i after this update slot is $\lceil s_i' \rceil$. Overwrite s_i with s_i' and return to Step 2 at the next update slot.

SINR is estimated and hence the uplink uses the same channel as in downlink.

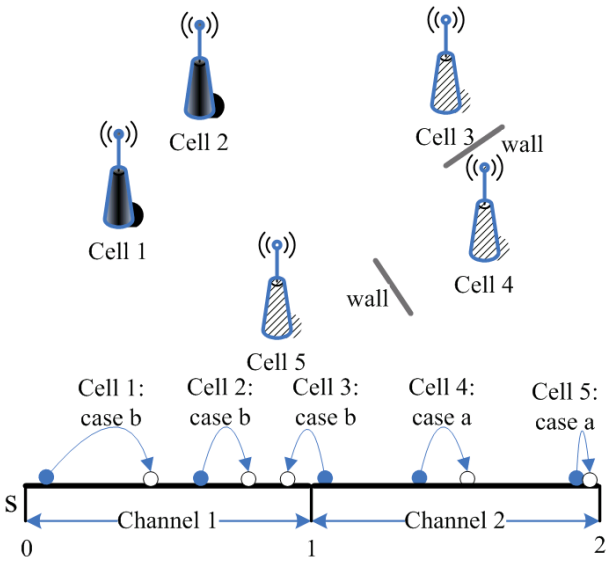


Fig. 3.9: An example of FiDCA with 5 cells.

An example of FiDCA is given in Fig. 3.9, where 5 cells exist in a network. Cell 1 and 2 are assigned on channel 1; cell 3, 4 and 5 are assigned on channel 2. The connections between cell 3 and cell 4, between cell 4 and cell 5 are blocked by walls, giving very high path loss values. The initial status for each cell can be seen from Fig. 3.9, and the path loss values between every two APs are summarized in Table 3.1.

Table 3.1: Path loss values [dB] between the APs.

AP#	1	2	3	4	5
1	–	60.2	74.2	73.6	64.0
2	60.2	–	70.0	72.4	67.2
3	74.2	70.0	–	74.8	65.9
4	73.6	72.4	74.8	–	80.3
5	64.0	67.2	65.9	80.3	–

According to the channel allocation pattern and the path loss values, cell 1, 2 and 3 are not satisfied with their channel allocation, and they would jump towards the other channel; cell 4 and 5 are satisfied, and they jump towards the right boundary. After the update, cell 3 successfully switches to the first channel. Next is cell 2, which will change to the second channel and will not interfere with cell 1 any longer. Therefore, cell 1 becomes satisfied with its channel allocation, and adjusts the indicator s_1 towards the left boundary. Finally, the system stabilizes

when cell 1 and cell 3 are assigned on one channel, and cell 2, 4, 5 are assigned on the other one. The evolution of the status indicator for each cell and the network channel allocation pattern is presented in Fig. 3.10, with $\beta = 0.4$ and $G = 5$ dB.

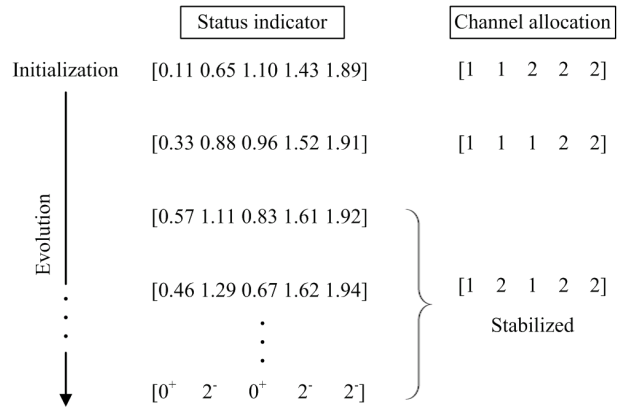


Fig. 3.10: Status indicator and channel allocation pattern versus operation time.

According to the descriptions above, the FiDCA algorithm operates in a fully distributed manner. However, it is also possible to be implemented in a centralized manner. For this purpose, each cell has to report the path loss values of its neighbors to the central controller. The central controller then estimates the channel quality and performs the iterations for each cell. It will stop when the network is converged, and then provide the channel assignment to each cell. By operating in the centralized manner, the convergence time is saved because the central controller can estimate the channel quality with each possible allocation pattern, without the need of actual transmission and measuring the SINR.

3.3.2 Channel-Aware Selfish Dynamic Spectrum Sharing

The FiDCA algorithm presented before aims to achieve the optimal reuse-2 pattern without centralized planning. In most cases it can significantly improve the performance. However, in cases with very low activity factors, i.e., many APs are not activated, the inter-cell interference is weak, and using only half of the total spectrum is a waste of resources. Also in cases with very dense AP deployment, reuse-2 may fail to offer enough protection against the surrounding cells. This motivates to use an intelligent algorithm that is aware of the interference condition, and can trade off between the performance of a cell and the interference it generates to the neighbors. For this purpose, a decentralized algorithm called Dynamic Spectrum Sharing with Selfishness (DS^3) is developed.

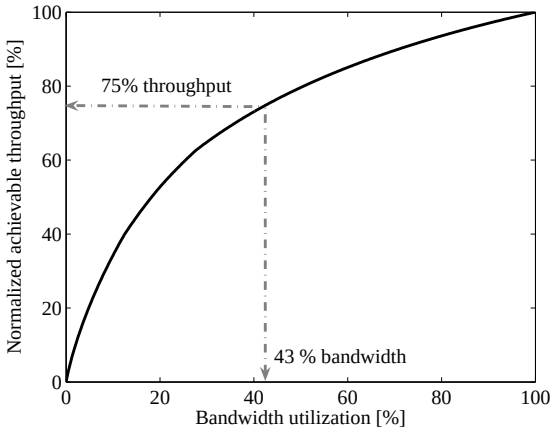


Fig. 3.11: Basic idea behind the DS^3 algorithm.

The fundamental principle behind DS^3 is that rather than using all possible spectrum to achieve the highest peak throughput at each moment, one should sacrifice own throughput by using less spectrum so as to generate lower interference to others. If every cell follows the same rule, the total network interference will be reduced and the overall performance may increase. Fig. 3.11 gives an example of the relationship between the bandwidth utilization and achievable throughput in the downlink transmission of one LTE-Advanced cell. It is generated from one snapshot of a small indoor home scenario, following the system settings in Table A.1. The PRBs are sorted so that the ones with high SINR are used before those with low SINR. As can be seen from this figure, the use of heavily interfered spectrum reduces the spectral efficiency, and hence the slope of the curve gets lower as the spectrum utilization increases. In this example, 43% of the spectrum can achieve 75% of the maximum throughput. Meanwhile, the interference generated from this cell is limited to a much narrower bandwidth, and the neighboring cells are expected to benefit from this by prioritizing the less interfered spectrum.

The DS^3 algorithm makes use of this finding that using only part of the available spectrum (with high SINR) can reduce the interference to other cells, without losing much of the performance for the current cell. It first estimates the maximum achievable throughput of a cell, based on the SINR measurement from the users. Afterwards, it uses only the spectrum with good channel quality to achieve X% of the maximum throughput, and releases the remaining spectrum. It can be realized using the flowchart shown in Fig. 3.12, and the steps are described below:

Step 1: Initialization. In this step, the new AP will randomly pick up PRBs according to its traffic requirement.

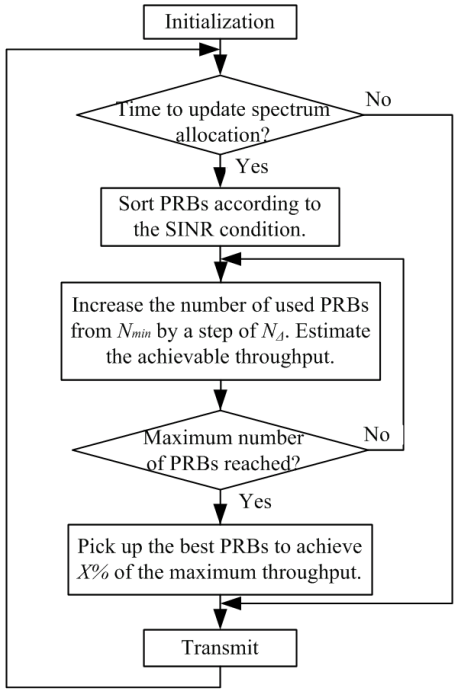


Fig. 3.12: Flowchart of DS^3.

Step 2: Obtain SINR measure and then sort the PRBs according to the average SINR values (over the multiple users) in the descending order. The SINR should be averaged over a certain time interval such that the effect of fast fading is eliminated.

Step 3: Increase the number of selected PRBs from N_{min} by a step of N_{Δ}, which is the incremental granularity, and estimate the achievable throughput. This step is repeated until the maximum number (N_{max}) of used PRBs is reached. In this step, a PRB with high SINR is always selected before the ones with low SINR. Note that here the AP just estimates the achievable throughput; it will select the preferable PRBs based on this estimation in the next step.

Step 4: Estimate the maximum achievable throughput and pick up the best PRBs that can offer $X\%$ of the maximum throughput, as shown in Fig. 3.11.

Step 5: Transmit with the resulting allocation pattern, and wait until its next update slot.

For the implementation of DS^3 in real systems, a few things are worth mentioning:

- The neighboring cells should take turn to update their spectrum allocation. If one AP is updating the spectrum allocation, its neighbors should follow their old allocation pattern so that the interference condition remains unchanged. This is ensured in the initialization phase by having an update queue of the APs. Coordination among cells is required to build this queue. If such coordination is unavailable, a random waiting window or a probability of updating could be applied in each cell. In this case, collision may happen between neighboring cells, giving degraded performance.

- The selfish factor X is used to prevent one cell from taking the whole spectrum. The higher the X, the more selfish is a cell.

- In Step 3, N_{min} is used to guarantee the basic service within a cell. Together with N_{max} and N_{Δ}, they can limit the searching space for DS^3 and trade off between complexity and performance. To maintain fairness, the value of X, N_{min} and N_{max} should be agreed by all cells as part of a policy. The granularity N_{Δ} can be different in different APs, depending on their processing speeds.

3.4 Performance of the Proposed Algorithms

In this section the performance of the two proposed FSU algorithms is evaluated and compared with the reference performance of frequency reuse. In the evaluation, $\beta = 0.1$ and $G = 20$ dB is chosen for FiDCA. The values of these two parameters mainly affect the convergence time but have nearly no effect on the system performance once converged. In the third step of DS^3, each cell is allowed to have a minimum of 1/3 of the total spectrum, i.e. $N_{min} = 167$. N_{max} is set to be the same as the total number of PRBs, and the same granularity step $N_{\Delta} = 2$ is adopted in all cells to control operation complexity. The downlink performance is considered at first. Later, the uplink performance is also provided, confirming the applicability of the algorithms in both transmission directions.

3.4.1 Downlink Performance of FiDCA

Fig. 3.13 shows the downlink performance of FiDCA in the home scenario with 16 cells and different activity factors. The performance of frequency reuse-2 is also plotted for reference, and the throughput values are normalized with respect to reuse-1. For all activity factors, FiDCA has a similar performance as reuse-2, meaning that by the dynamic allocation, the strongest interference is successfully avoided, and the interference is efficiently reduced.

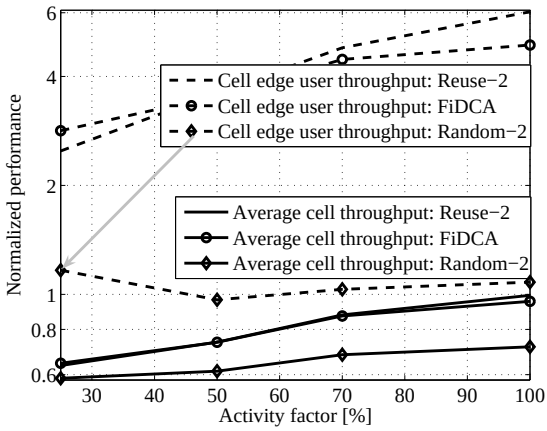


Fig. 3.13: Normalized performance of Reuse-2 and FiDCA with respect to Reuse-1. Evaluation is performed in the downlink of a 16-cell home scenario.

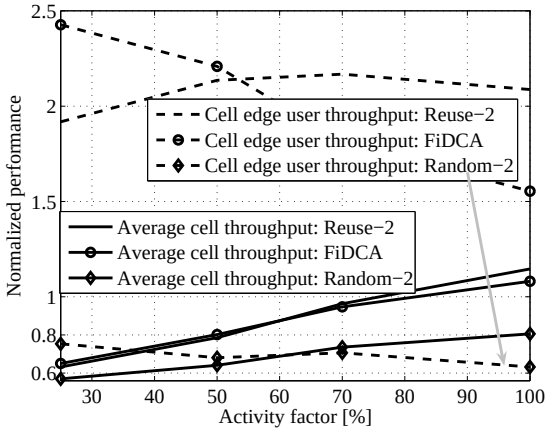


Fig. 3.14: Normalized performance of Reuse-2 and FiDCA with respect to Reuse-1. Evaluation is performed in the downlink of a 16-cell office scenario.

In the office scenario, because of the fixed AP position, the gain of FSU algorithms as compared to reuse-1 is much reduced. The cell edge user throughput of FiDCA is presented in Fig. 3.14, which is 1.6∼2.5 times the performance of reuse-1. With very low activity factors, FiDCA gives better performance than reuse-2 in both average cell throughput and cell edge user throughput. When more cells are active, it still achieves relatively good performance as compared to frequency reuse.

FiDCA is an evolutionary algorithm that starts from random channel allocation

and evolves towards planned frequency reuse. The performance of random reuse-2 is also presented in Fig. 3.13 and Fig. 3.14, which clearly shows the benefit of the evolutionary process using FiDCA.

3.4.2 Downlink Performance of DS^3

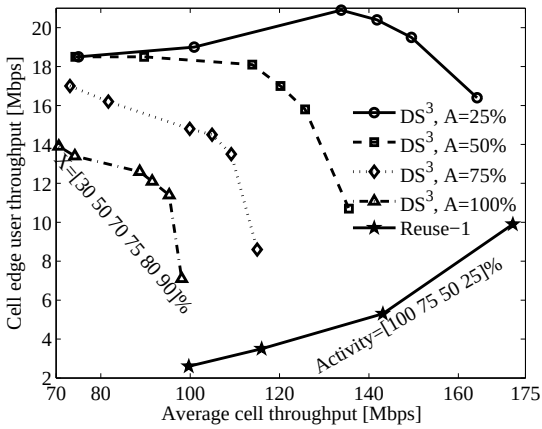


Fig. 3.15: Performance of DS^3 with different selfish factors and cell activity factors in the downlink of a 16-cell home scenario, with reuse-1 as the reference.

For DS^3, the selfish factor X is essential for the performance. Various selfish factors have been tested, and the results in the downlink transmission of an extended home scenarios with different activity factors are summarized in Fig. 3.15, with reuse-1 being the reference. In the legends of this figure, 'A' is short for 'activity factor'. The following observations can be made from this figure:

1. A very small selfish factor X is inefficient because it causes very low spectrum utilization, which cannot be compensated by the interference reduction.
2. A high selfish factor of 90% achieves similar average cell throughput as reuse-1, but still benefits in cell edge user throughput. However, the cell edge user throughput drops dramatically for very high selfish factors. An even higher value of X tends to use the whole spectrum, and hence similar performance as reuse-1 is expected.
3. When the activity factor is high, the loss due to a high value of X is more significant than the case with low activity factor. On the other hand, a very small X gives relatively worse performance when the activity factor is low. However, the best trade-off is obtained by having X lying between 70∼80%, which is the same for all simulated activity factors.

In the following, a selfish factor of 75% is used for evaluating the performance of DS^3. The performance is presented in Fig. 3.16 and Fig. 3.17 for the home and office scenarios, respectively. Frequency reuse-2 is plotted as reference, and the performance is normalized with respect to reuse-1.

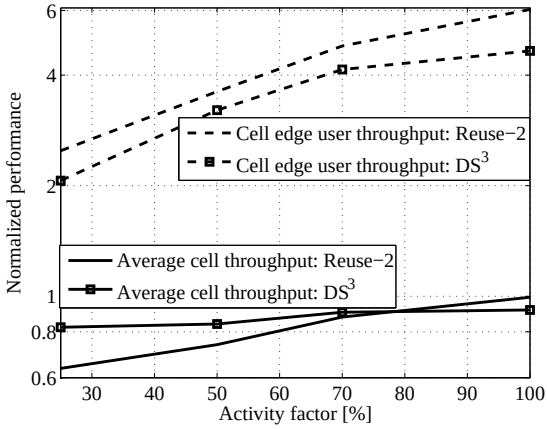


Fig. 3.16: Normalized performance of Reuse-2 and DS^3 with respect to Reuse-1. The evaluation is performed in the downlink of a 16-cell home scenario.

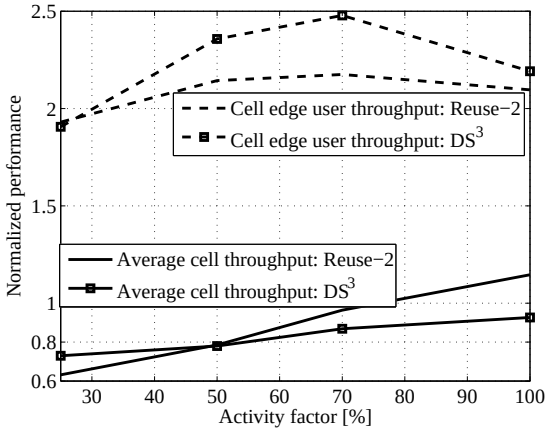


Fig. 3.17: Normalized performance of Reuse-2 and DS^3 with respect to Reuse-1. The evaluation is performed in the downlink of a 16-cell office scenario.

Generally speaking, DS^3 achieves similar cell edge user throughput as reuse-2, which is much better than reuse-1. In terms of average cell throughput, the cho-

sen selfish value of $X = 75$ offers better performance than reuse-2 at low activity factors because of its high spectrum utilization. In this situation, the SINR is very high, and the throughput is roughly proportional to the bandwidth. However, when the activity factor is high (more cells are active), the inter-cell interference is high. As a consequence, DS^3 with the same selfish factor of $X = 75$ has lower spectrum utilization than in the case with sparse number of active cells. This guarantees good cell edge user throughput, but degrades the average cell throughput, as compared to reuse-2. A high value of X could potentially be used, which will improve the average cell throughput but at the same time reduce the cell edge user throughput (as shown in the left-most curve of Fig. 3.15).

3.4.3 Uplink Performance of FiDCA and DS^3

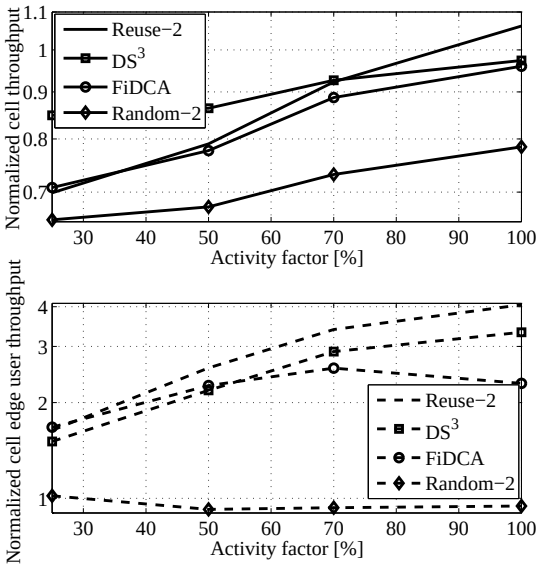


Fig. 3.18: Normalized performance of Reuse-2, DS^3 and FiDCA with respect to Reuse-1. The evaluation is performed in the uplink of a 16-cell home scenario.

The downlink channel quality measurement has been used in this chapter to evaluate the uplink performance of FiDCA and DS^3. Therefore, the uplink performance is not fully optimized according to the experienced channel condition. Even with this imperfection, Fig. 3.18 and Fig. 3.19 show that these two algorithms offer much higher cell edge user throughput than reuse-1. Their average cell throughput is also better than or similar as reuse-1 at high activity factors. When the activity factor is low, both reuse-2 and FiDCA have worse average cell throughput than

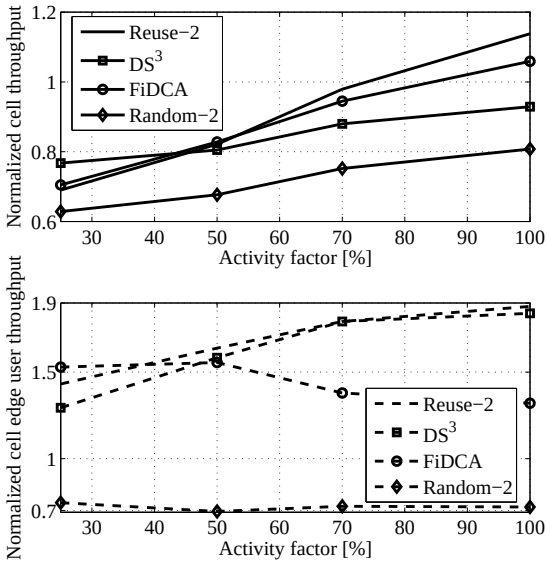


Fig. 3.19: Normalized performance of Reuse-2, DS^3 and FiDCA with respect to Reuse-1. The evaluation is performed in the uplink of a 16-cell office scenario.

reuse-1. However, DS^3 achieves around 80% the performance of reuse-1, which is better than the other algorithms. If these algorithms are to be implemented in real systems, where the uplink channel measurement is available, a better uplink performance than presented here is expected. As compared to random channel selection, their is also a huge gain in the uplink by using the developed algorithms.

3.4.4 Comparison between FiDCA and DS^3

Both FiDCA and DS^3 are designed to facilitate the flexible spectrum sharing of the local area networks in a self-organized manner. DS^3 is more flexible than FiDCA because it allows a cells to use different amount of spectrum by tuning the selfish factor X. A properly chosen X ensures sufficient spectrum usage with sparse number of active cells, as well as offers enough inter-cell interference reduction when more cells are active. FiDCA has a fixed spectrum utilization of 50% (reuse-2 alike), it is attractive when cell density is neither very low nor very high. For very low cell density, it causes inefficient spectrum usage because the inter-cell interference is very low. For very high cell density, the fixed utilization may fail to offer enough reduction of inter-cell interference.

Despite the lack of flexibility with respect to cell density, FiDCA is also advantageous in the following aspects: first, it operates based on the channel measurement within each cell, and there is no information exchange or coordination among neighboring cells. DS^3, on the other hand, requires a common queue to control the spectrum updating of each cell. Information exchange among neighboring cells is needed to set up this queue. Secondly, FiDCA is fairly simple. It divides the total spectrum into two channel segments, and operates on the channel segment level. The complexity is very low compared to DS^3 that operates on the PRB level and has to estimate the achievable throughput using different number of PRBs. For example, with $N_{min} = 167$, $N_{max} = 500$ and $N_{\Delta} = 2$, DS^3 has to estimate 166 possibilities before deciding which PRBs to use, whereas FiDCA just needs to adjust the status indicator s_i according to (3.2), and updates the channel allocation as $\lceil s_i \rceil$. Thirdly, FiDCA can be used in both local area and wide area networks, but DS^3 is applicable only for the local area networks with low transmit power from the users. The reason is that the selection of PRBs in DS^3 is based on the channel quality, and hence non-consecutive PRBs may be selected by a cell. In the wide area networks, the users are transmitting with high power; they will suffer from high Peak to Average Power Ratio (PAPR) if scheduled on non-consecutive PRBs [93]. PAPR is not a critical issue for the local area network, because the users are transmitting with very low power, as can be seen from Fig. 2.3.

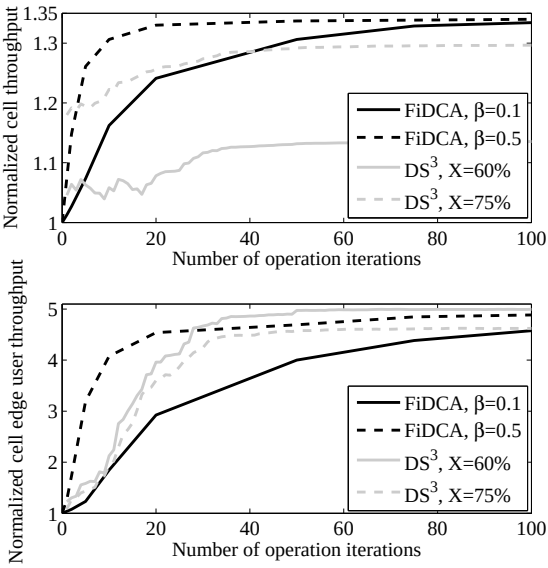


Fig. 3.20: Convergence time for FiDCA and DS^3 in the downlink of a 16-cell home scenario. All APs are activated.

Both FiDCA and DS^3 are iterative algorithms and require a certain time for the

performance to converge. Fig. 3.20 illustrates the convergence time for them in the downlink of an extended home scenario. The performance is normalized with respect to random channel allocation (FiDCA with 0 iteration). From this figure it can be seen that, with 16 active cells, the performance of DS^3 is stabilized after around 30 iterations, and this is insensitive to the selfish value. The time required by FiDCA to reach the optimal pattern depends on the jumping step α, which is ultimately decided by the values of β and G. While a small jumping step requires long convergence time, a very large jumping step may break the equilibrium, and the system may never converge. By properly tuning the values of β and G, the best jumping step to minimize the convergence time can be found, and it is similar as the time required by DS^3.

3.5 Conclusion

This chapter has studied several possibilities of improving the spectrum utilization efficiency in the local area networks. The performance of frequency reuse is investigated in the beginning. It is observed that the system performance can be significantly improved when using a reuse factor of 2, as compared to universal frequency reuse (reuse-1). However, frequency reuse requires network planning or a central controller to assign frequency channels properly to the cells. This is usually unfeasible in local area networks. It is therefore used only to generate the reference performance.

Motivated by the superior performance of frequency reuse-2, and being aware of its limitations, several FSU algorithms have been developed to improve the system performance in the distributed and self-organized manner. One of the algorithms is FiDCA, it aims to achieve the optimal reuse-2 pattern by allowing each cell to jump towards the channel with better quality than the other. Another algorithm, the DS^3, reduces inter-cell interference by preventing cells from operating on spectrum with poor SINR. FiDCA is simple to implement, and it offers much better cell edge user throughput than reuse-1. However, it is not adaptive to the cell density, and hence the relative performance is poor when cells are sparsely deployed. DS^3 offers the flexibility to trade off between average cell throughput and cell edge throughput. By tuning the selfish factor, it allows a cell to use more spectrum when inter-cell interference is low, or less spectrum if the interference is high. The selection between FiDCA and DS^3 in a specific local area network can be done depending on the variation range of the cell density and the processing speed of the AP.

Chapter 4

Carrier Load Balancing and Packet Scheduling for Wide Area Multi-CC Systems

4.1 Introduction

This chapter focuses on resource allocation for Long Term Evolution (LTE)-Advanced systems with aggregation of multiple Component Carrier (CC)s, i.e., how to assign the CCs to each user, and how to multiplex users in each CC. The gain of using carrier aggregation over independent carriers will also be shown.

The LTE-Advanced systems are required to be backwards compatible, meaning that users in such systems may have access to only one, or multiple CCs, depending on their terminal types and the Quality of Service (QoS) requirements. E.g., an LTE-Advanced user with voice traffic can be assigned a single CC, while still being able to fulfill the QoS requirement; an LTE Rel'8 user is restricted to operate on one CC, despite the QoS requirement. In this chapter, the LTE-Advanced users are admitted on all CCs to maximize the multi-user diversity. The possibility of assigning less number of CCs to the LTE-Advanced users will be analyzed in the next chapter, where the purpose is to reduce control channel and uplink feedback overhead as well as the user power consumption without losing much of the downlink performance.

Having users operating on a subset of the available CCs, it is desirable to have approximately equal load on different CCs, and hence carrier load balancing should be used. Several known approaches for the load balancing across CCs are studied in this chapter and the best one is identified. The existence of both single and multi-CC capable users also gives rise to challenges in maintaining fairness among the users along with system efficiency, which should be properly handled by the packet scheduler. A simple, but efficient cross-CC packet scheduler is proposed. As compared to independent scheduling per CC, it maximizes the network utility and improves the fairness. To comply with the backwards compatibility, the 3GPP has decided to use independent layer-1 transmission, which contains link adaptation and HARQ etc, per CC in coherence with the LTE-Rel'8 assumptions [18].

With the system setup described in Fig. 1.2, various layer-3 carrier load balancing methods and layer-2 scheduling algorithms are studied. The target is to optimize the resource allocation with the existence of multiple CCs. The study is conducted both in the form of a simple full buffer model with a constant number of users per cell, and a more realistic finite buffer model with Poisson arrival. The Orthogonal Frequency Division Multiple Access (OFDMA) based LTE-Advanced downlink is chosen as a case study. The LTE-Advanced resource allocation process addressed in this chapter has similarities with related problems for other systems such as multi-carrier HSDPA [51, 52] and multi-carrier CDMA [53–58]. However, these studies do not consider the case when users can access different number of CCs. Also, the CDMA based systems lack the flexibility to multiplex users in frequency domain.

The rest of this chapter is organized as follows: Section 4.2 outlines the different load balancing and Packet Scheduling (PS) techniques under investigation. In Section 4.3 both theoretical analysis and simulation results for load balancing are presented, with independent scheduling and full buffer traffic model. The performance of different PS algorithms is compared with each other in Section 4.4, using both full and finite buffer traffic models. Finally, Section 4.5 concludes the chapter.

4.2 Radio Resource Management

The Radio Resource Management (RRM) framework for LTE-Advanced is similar to that of LTE-Rel'8, as discussed in Section 1.2.2. However, a new functionality is introduced with LTE-Advanced, which is referred to as CC assignment. It configures a set of CCs for each user with Radio Resource Control signaling. The user may afterwards be scheduled on this CC set. In this chapter, the LTE-Advanced users are admitted on all CCs to maximize the multi-user diversity. The base station therefore only needs to select a proper CC for each of the Rel'8 users, using

CC selection, as shown in Fig. 4.1. With multiple users on each CC, frequency domain channel aware PS can exploit the frequency and user domain diversity to improve the system performance compared with channel blind scheduling [94–97]. PS is performed for all users within each CC, but the scheduling metric can be calculated based on the statistics across the CCs. Independent layer-1 transmission is performed on each CC after the packet scheduling [18]. This means that the transmission block is coded separately within each CC, and therefore, the physical layer transmission cannot benefit from the frequency diversity across the CCs.

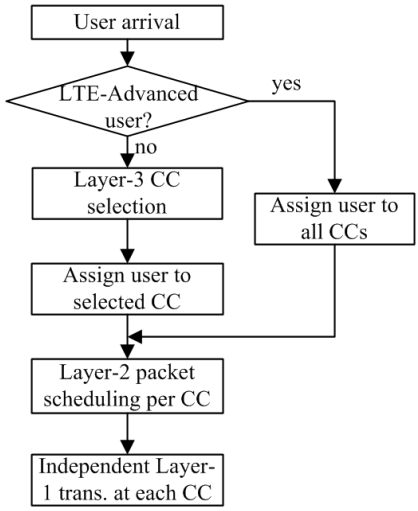


Fig. 4.1: Different ways to treat the LTE-Advanced users and the LTE-Rel'8 users in a multi-CC LTE-Advanced system.

The CC assignment functionality and the PS are the most important parts of the RRM in the base station. The CC assignment is based on the load condition, which is available locally at the base station. Meanwhile, to perform channel aware PS, the feedback of Channel Quality Information (CQI) in the uplink direction is required. The decisions of the RRM are used locally at the base station, as well as sent through the downlink control channel to the users.

The model of a multi-CC system with channel aware PS is depicted in Fig. 4.2, where the number of users in the system is K, and N CCs are aggregated to form the wide bandwidth. The aggregated CCs can have different bandwidth, and hence different number of Physical Resource Block (PRB)s. $M(n)$ denotes the available PRBs in each CC. Depending on the terminal category, one user is able to access either the full set or a subset of the CCs, denoted by $C(k) \subset \{1, 2, \ldots, N\}$. $C(k)$ is the output of the CC assignment function. Scheduling is performed after the CC assignment. It decides the PRBs each user should operate on. For user k on the

n^{th} CC at time t, $S(k,n,t) \subset M(n)$ is used to denote the assigned PRBs. Because each PRB is assigned to only one user, a direct mapping exists from the m^{th} PRB on the n^{th} CC to the user with index k: $k = P(m,n,t)$. Hence, $P(\cdot)$ can be seen as the output of the PS. This multi-CC model will be used later to prove that the proposed cross-CC scheduler maximizes the network utility.

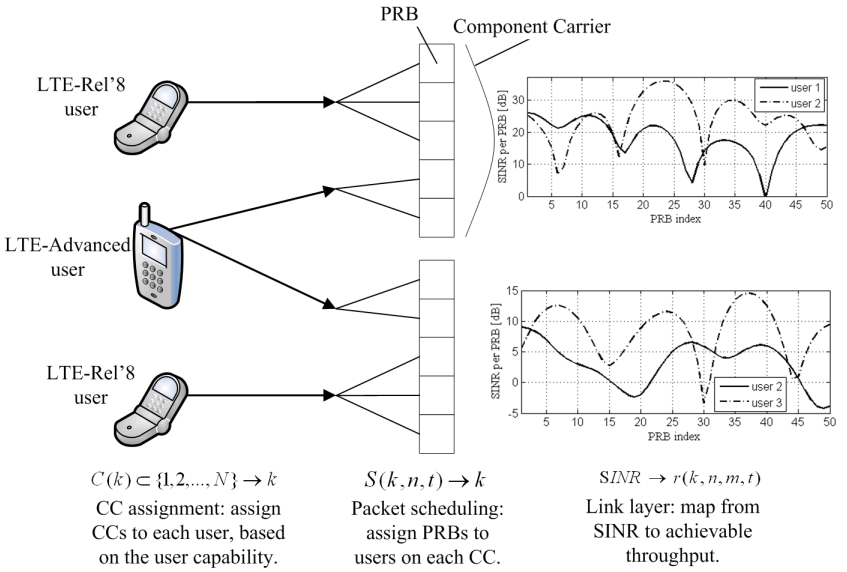


Fig. 4.2: System model of a multi-CC system with channel aware scheduling.

4.2.1 Carrier Load Balancing Methods

As mentioned before, the study here is based on the assumption that the LTE-Advanced users are assigned on all CCs. Two methods for layer-3 CC load balancing are studied for the LTE-Rel'8 users, which are described in the following. The assigning of CCs is performed for each user upon arrival, and the assigned CCs remain unchanged during the user's transmission session. Note that the case when all users are LTE-Advanced is also known as the 'carrier aggregation' mode in literature. The case when all users are Rel'8 is referred to as the '$N\times$independent (or single)-carrier' mode [51, 98–100].

4.2.1.1 Least Load Balancing

The Least Load (LL) balancing [53] is also referred to as Combined Carrier Channel Assignment [55]. The basic principle is to assign the newly arrived user to the CC that has the lowest number of users. Thus, it tries to distribute the load evenly to all CCs. However, there might be small load variations on different CCs as the number of Rel'8 users does not always divide equally on the number of CCs, or due to users randomly terminating their connections. With the full buffer traffic model, LL is also equivalent to the Round Robin (RR) balancing in [56].

4.2.1.2 Mobile Hashing Balancing

The Mobile Hashing (MH) balancing method [56], also known as the Independent Carrier Channel Assignment [55] or Random Carrier [53], relies on the output from the terminal's hashing algorithm. The output hash values are uniformly distributed among a finite set, which maps directly on the CC indices [101]. Therefore, MH basically corresponds to random CC selection, which provides balanced load across CCs in the long term. However, at each time-instant, the load across CCs is not guaranteed to be balanced. The unbalanced cell load will lead to a lower multi-user gain than LL, as proved in the following.

Let K_n denote the number of users on the n^{th} CC. Assuming the CCs have equal bandwidth, the number of users with LL balancing is $\tilde{K} = \sum_{n=1}^{N} K_n/N$ per CC. It is known that

$$\prod_{n=1}^{N} K_n \leq \tilde{K}^N \tag{4.1}$$

The equality holds when $K_n = \tilde{K},\ \forall\ 1 \leq n \leq N$. Take the logarithm of (4.1), the following is obtained:

$$\sum_{n=1}^{N} \ln K_n \leq N \ln \tilde{K} \tag{4.2}$$

When channel aware packet scheduling is used, the multi-user scheduling gain follows approximately a logarithmic function of the number of users, in the form of $G(K) = A \ln K + B$ (see Section B.5). Therefore, (4.2) is equivalent to:

$$\sum_{n=1}^{N} G(K_n) \leq N \ln G(\tilde{K}) \tag{4.3}$$

Equation (4.3) means that the multi-user diversity is maximized when load is balanced across the CCs. Therefore, when channel aware packet scheduling is used, LL will have better average cell throughput than MH.

4.2.2 Packet Scheduling

The rule of PS is to facilitate the sharing of the time-frequency resource among the multiple users within each TTI. When the number of users is high, it can be a very heavy calculation burden to acquire and process the channel information from the users, and to come up with the resource allocation for each of them. To reduce complexity, the scheduling is usually broken down to two steps: first, time domain scheduling, which selects a subset of the available users for resource allocation; secondly, frequency domain scheduling, which allocates the frequency domain resources to the users that pass the time domain scheduler. It exploits the multi-user diversity to improve the system performance based on the channel quality feedback [102].

In this study, an RR scheduler is used in time domain, which passes maximum 20 users to the frequency domain. RR is channel blind, therefore its scheduling metric is decoupled from the one used in frequency domain [103]. In the frequency domain, a channel aware Proportional Fair (PF) scheduler [97,104] is selected to exploit the frequency domain diversity. It operates independently on each CC, and is hereafter named independent PF. However, independent PF cannot efficiently handle the case when users are assigned different number of CCs. An extension to it is proposed, which takes the user throughput across the CCs into consideration, and improves the fairness among users. It will also be proved that the proposed cross-CC scheduler maximizes the network utility. These two schedulers are described below.

4.2.2.1 Independent PF Scheduling

With the independent PF scheduler, the resource is assigned to the user that maximizes the following scheduling metric [104]:

$$\mathcal{M}_{k,n,m,t} = \frac{r(k,n,m,t)}{R(k,n,t)} \tag{4.4}$$

In (4.4), $r(k,n,m,t)$ is the expected throughput for user k on the n^{th} CC and the m^{th} PRB at time instance t. It is estimated based on the user channel quality (typically a quantized SINR report) and the link level capacity of the system, including the effect of modulation and coding rate adaptation, MIMO rank adaptation, etc. $R(k,n,t)$ is the average throughput for user k on the n^{th} CC up to time t.

It is shown in [105] that with sufficient large value of t, $R(k,n,t)$ weakly converges to a constant value for a certain user k. This assumption of stationary throughput is also used in [106]. In reality, an average window is used for the averaging,

instead of counting all the past throughput values [106]. On the n^{th} CC at time $t+1$, $R(k,n,t)$ is updated according to the following equation [104]:

$$R(k,n,t+1) = R(k,n,t)(1-1/T) + 1/T \sum_{m \in S(k,n,t)} r(k,n,m,t) \tag{4.5}$$

where T is the length of the sliding average window for user throughput measurement. The value of T should be chosen to offer a good estimation of the average throughput, with the ability to track changes in the channel characteristics [105] (e.g. 100 slots in [96], 1000 slots in [104]).

The traditional PF guarantees allocation fairness within each CC, but disregards the fact that users are assigned with different number of CCs. Therefore it cannot achieve the global fairness. E.g., consider two users with the same average channel quality and fast fading statistics, one is LTE-Advanced and the other is Rel'8. The PF scheduler gives an equal share of resources to the users on the CC that the Rel'8 user is served [106]. However, the LTE-Advanced user is also scheduled on other CCs. Therefore it gets much more resources than the Rel'8 user.

4.2.2.2 Cross-CC PF Scheduling

By taking the statistics from all CCs into consideration, the scheduler can achieve an overall better resource allocation than independent PF. In order to reduce the complexity for upgrading the existing LTE systems, the proposed scheduler still operates within each CC. The only difference from independent scheduling is that it takes the aggregated user throughput over all CCs into account, i.e., $R(k,t) = \sum_{n \in C(k)} R(k,n,t)$. The scheduling metric becomes:

$$\mathcal{M}'_{k,n,m,t} = \frac{r(k,n,m,t)}{R(k,t)} \tag{4.6}$$

With (4.6), the LTE-Advanced users have reduced scheduling metrics because their overall throughput is higher than the throughput per CC. On the other hand, the LTE-Rel'8 users maintain their scheduling metrics, because their transmissions and receptions are restricted on only one CC. They are thereby prioritized as compared with the LTE-Advanced users in resource allocation, which meets the objective of improving fairness among users.

The only requirement for upgrading from independent scheduling is to aggregate the past user throughput across all CCs. Because the throughput within each CC is known at the base station, there is just one extra sum operation for the scheduling in all CCs. No further information exchange or collaboration is needed for the schedulers in different CCs. Thus, it imposes only marginal calculation

complexity on top of the independent PF.[1]

4.2.2.3 Utility Maximization of Cross-CC PF

One mathematical model of the network utility is provided here. Based on this model, it will be proved that the cross-CC PF maximizes the utility.

Let U denote the network utility, defined as

$$U(t) = \sum_{k=1}^{K} \ln R(k,t) \tag{4.7}$$

According to [105, 107], the problem of maximizing the utility is equivalent to maximizing the increase of $U(t+1)$ over $U(t)$, which is:

$$\Delta U = U(t+1) - U(t) = \sum_{k=1}^{K} \big(\ln R(k,t+1) - \ln R(k,t) \big) \tag{4.8}$$

The large value of T implies that the change in average user throughput during each scheduling interval is rather small. Therefore, the first order Taylor series expansion of $\ln R(k,t+1)$ at $x_0 = R(k,t)$ can be applied: $\ln R(k,t+1) \approx \ln R(k,t) + \frac{R(k,t+1)-R(k,t)}{R(k,t)}$. Equation (4.8) then becomes:

$$\begin{aligned}
\Delta U &\approx \sum_{k=1}^{K} \frac{R(k,t+1) - R(k,t)}{R(k,t)} \\
&= \sum_{k=1}^{K} \frac{\sum\limits_{n \in C(k)} \left(R(k,n,t)(1-1/T) + 1/T \sum\limits_{m \in S(k,n,t)} r(k,n,m,t) - R(k,n,t) \right)}{R(k,t)} \\
&= -K/T + 1/T \sum_{k=1}^{K} \sum_{n \in C(k)} \sum_{m \in S(k,n,t)} \frac{r(k,n,m,t)}{R(k,t)}
\end{aligned} \tag{4.9}$$

Because each PRB is assigned to only one user, the summation over all users on its accessible CCs and the assigned PRBs on these CCs is equivalent to the summation over all PRBs, i.e., $\bigcup\limits_{1 \le k \le K} S(k,n,t) = M(n)$ for each CC, with $S(k,n,t) = \emptyset$ if the

[1] The cross-CC scheduling concept also works for the Generalized Proportional Fair (GPF) scheduler [96]. GPF can prioritize users with different channel qualities. It offers the trade-off between average cell/user throughput and cell edge user throughput when the full buffer model is used. With the finite buffer model, GPF can improve the system performance by favoring the users with good channel qualities, and letting them depart from the system in a fast manner. Detailed performance analysis of the GPF scheduler is summarized in Section D.1.

user is unscheduled. Therefore, (4.9) can be rewritten as:

$$\Delta U \approx -K/T + 1/T \sum_{n=1}^{N} \sum_{m \in M(n)} \frac{r(k,n,m,t)}{R(k,t)} \tag{4.10}$$

Both K and T are fixed values, therefore different values of ΔU can be achieved by assigning the resources to different users. It is maximized by allocating the m^{th} PRB on the n^{th} CC at time t to a user k' that satisfies:

$$k' = P(m,n,t) = \underset{k}{argmax} \frac{r(k,n,m,t)}{R(k,t)} = \underset{k}{argmax} \mathcal{M}'_{k,n,m,t} \tag{4.11}$$

Equation (4.11) makes use of the scheduling metric in (4.6), therefore indicates that the cross-CC PF scheduler maximizes the network utility.

4.3 Performance With Different Carrier Load Balancing Methods

In this section, the performance of different carrier load balancing methods is analyzed both theoretically and via simulation. The analysis is carried out using independent scheduling per CC and the full buffer traffic model. The performance with cross-CC PF will later be evaluated based on simulations. The analysis for the finite buffer model is put in Appendix C, due to the complexity in modeling the dynamic arrival/departure of the users. The simulation methodology and assumptions are summarized in Appendix A.2.

4.3.1 Theoretical Analysis for LL and MH

Let us consider a general case with N aggregated CCs and K active users per cell. If each user has probability α of being an LTE-Advanced user, then the probability of having K_α LTE-Advanced users out of the K active users equals

$$P_{K_\alpha} = \binom{K}{K_\alpha} \alpha^{K_\alpha} (1-\alpha)^{K-K_\alpha} \tag{4.12}$$

If the LL balancing is used, the average number of users on each CC equals

$$\tilde{K} = K_\alpha + (K - K_\alpha)/N \tag{4.13}$$

The average cell throughput with K_α LTE-Advanced users is

$$R_{cell,LL}(K_\alpha) = CG(\tilde{K})/G(\infty) \tag{4.14}$$

where $G(\tilde{K})$ is the multi-user gain that arises from the frequency domain packet scheduling; $G(\infty)$ is the maximum multi-user gain. C is the average cell throughput when the multi-user gain saturates. Combining (4.12), (4.13), (4.14), the average cell throughput with LL balancing is:

$$\tilde{R}_{cell,LL} = \frac{C}{G(\infty)} \sum_{K_\alpha=0}^{K} P_{K_\alpha} G(\tilde{K}) \tag{4.15}$$

If MH is used for carrier load balancing, each LTE-Rel'8 user has equal probability of being assigned any of the CCs, which is $1/N$. With K_α LTE-Advanced users, the probability for one CC to have k LTE-Rel'8 users is

$$P_k = \binom{K-K_\alpha}{k} \left(\frac{1}{N}\right)^k \left(\frac{N-1}{N}\right)^{K-K_\alpha-k} \tag{4.16}$$

The cell throughput with K_α LTE-Advanced users is

$$R_{cell,MH}(K_\alpha) = \frac{C}{G(\infty)} \sum_{k=0}^{K-K_\alpha} P_k G(K_\alpha + k) \tag{4.17}$$

and therefore the overall cell throughput equals

$$\tilde{R}_{cell,MH} = \frac{C}{G(\infty)} \sum_{K_\alpha=0}^{K} P_{K_\alpha} \sum_{k=0}^{K-K_\alpha} P_k G(K_\alpha + k) \tag{4.18}$$

4.3.2 Performance Comparison between LL and MH

The relative gain in average cell throughput by using LL over MH is shown in Fig. 4.3, with different ratios of LTE-Advanced users. The corresponding cell edge user throughput of the two balancing algorithms is presented in Fig. 4.4. From these two figures, the following observations are extracted:

1. In terms of average cell throughput, there is a good match between the theoretical estimates and the simulations, with maximum 1% deviation between the two. The theoretical expressions over-estimate the performance of LL balancing as compared to MH, due to the fact that the number of LTE-Rel'8 users may not be equally divided among the CCs. This is ignored in the theoretical analysis. When all users are Rel'8, LL balancing provides ∼7% higher cell throughput compared to MH. However, the gain decreases fast with the ratio of LTE-Advanced users, and it vanishes for more than 20% of the users are LTE-Advanced. The reason is that the assignment of CCs to the LTE-Advanced users is always balanced, and a higher ratio of LTE-Advanced users thus improves the overall balancing.

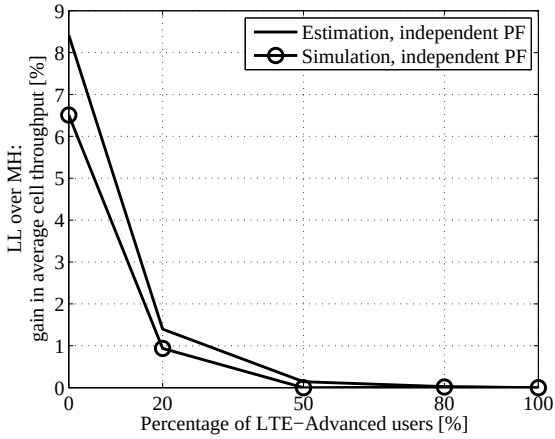


Fig. 4.3: Gain in average cell throughput by using LL balancing as compared to MH. Performance is evaluated with 10 users per cell and different ratios of LTE-Advanced users. Both estimated and simulated results are shown.

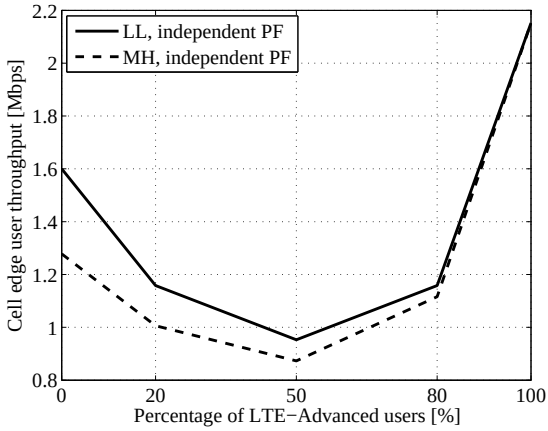


Fig. 4.4: Cell edge user throughput for LL and MH. Results are obtained via simulation with 10 users per cell and different ratios of LTE-Advanced users.

2. In terms of cell edge user throughput, LL is also more beneficial than MH, with maximum 25% gain when all users are Rel'8. As the ratio of LTE-Advanced users increases, the cell edge user throughput will first decrease, because less resources are available for each Rel'8 user. Meanwhile, the Rel'8 users will have less and less impact on the statistic of the user throughput. When the ratio is high enough for the LTE-Advanced users to dominate the throughput statistic, an increase of the cell edge user throughput with the

ratio will be experienced.

3. With the co-existence of both LTE-Advanced and Rel'8 users, it is difficult to estimate the cell edge user throughput. However, such an estimation is possible when all users are LTE-Advanced (the 'carrier aggregation' mode), or all are Rel'8 (the '$N\times$independent carrier' mode). With LL balancing the cell edge user throughput, when all users are Rel'8 compared to the case when all are LTE-Advanced, can be estimated using (B.22), as:

$$\frac{1/3\,(0.15\ln 3 + 1.05)}{4/10\,(0.15\ln 10 + 1.05)} = 73\% \tag{4.19}$$

In (4.19), the scaling factor of 1/3 in the numerator is due to the fact that when 10 Rel'8 users exist in a 4-CC system, some CCs will serve 3 users. On average, these users have much poorer throughput than the other ones, and will determine the cell edge user throughput. Similarly, there is a factor of 4/10 in the denominator because the 4 CCs are equally shared by the 10 LTE-Advanced users. This estimated loss is close to the simulation result, which is 74%.

It is also worth noting that all resources are utilized when there is at least one user per CC. There will be unused resources only when some CCs are empty (e.g. due to a low number of users and high ratio of LTE-Rel'8 users). With the given system configuration and the full buffer model, there is a sufficient number of users and hence the resources are fully utilized.

4.4 Performance With Different Packet Scheduling Algorithms

In Section 4.3 it has been shown that the Rel'8 users suffer from much lower performance than LTE-Advanced users, which causes degraded cell edge user throughput. This section makes use of the cross-CC PF as introduced in Section 4.2. It increases the scheduling priority for the Rel'8 users, and is expected to achieve higher cell edge user throughput than independent PF per CC.

4.4.1 Performance With Full Buffer Model

Fig. 4.5 shows the average cell throughput with independent or cross-CC scheduling. It can be seen that there is no obvious gain, or loss, by using cross-CC PF over independent PF. The performance when all users are Rel'8 is 11$\sim$12% worse

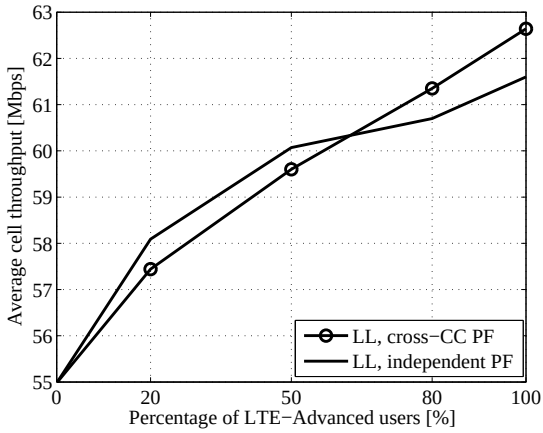


Fig. 4.5: Average cell throughput for independent or cross-CC PF with LL balancing. The full buffer traffic model is assumed, with 10 users per cell.

than the case when all are LTE-Advanced. This matches the estimation according to (B.21), which is $1 - (0.11 \ln 2.5 + 1.10)/(0.11 \ln 10 + 1.10) = 11\%$.

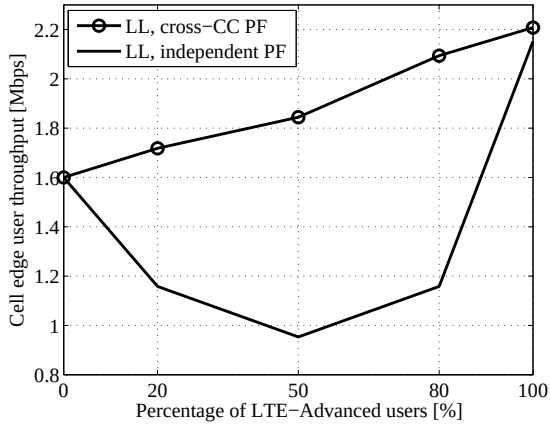


Fig. 4.6: Cell edge user throughput for independent or cross-CC PF with LL balancing. The full buffer traffic model is assumed, with 10 users per cell.

For cell edge user throughput, Fig. 4.6 shows that cross-CC PF has a significant improvement over independent PF. It improves the fairness between LTE-Advanced and Rel'8 users, and hence the cell edge user throughput constantly increases with the ratio. The maximum gain of cross-CC scheduling over independent scheduling happens when half of the users are LTE-Advanced, which is 90%. This clearly

shows the importance of using cross-CC PF for achieving acceptable cell edge user throughput. However, these two schedulers have similar performance when users can access the same number of CCs, e.g., all are LTE-Advanced or all are Rel'8.

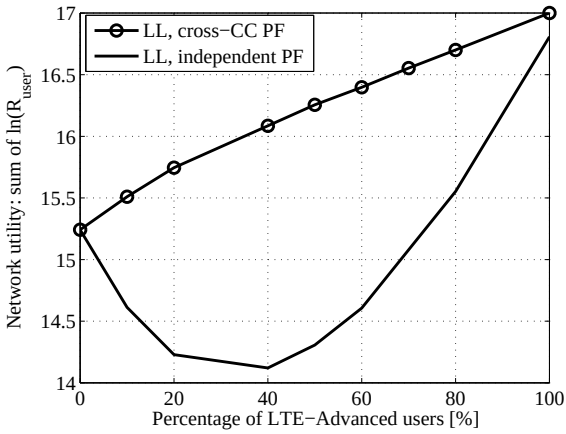


Fig. 4.7: Network utility with different ratios of the LTE-Advanced users. LL balancing is used, with 10 full buffer users per cell.

Fig. 4.7 provides the network utility according to the expression in (4.7). From this figure it can be seen that with a mixture of both LTE-Rel'8 and LTE-Advanced users, the independent PF suffers from much lower utility than cross-CC PF. The cross-CC PF achieves a higher utility as the ratio of LTE-Advanced users increases, because more users will exist in each CC, and hence there is more multi-user diversity for the scheduler to exploit. The independent PF sees a decrease in utility, which is due to the same reason as the poor cell edge user throughput, that it is unfair to the different kinds of users.

The CDF of the throughput for different kinds of users is plotted in Fig. 4.8. It clearly shows that the independent PF favors the multi-CC users by a factor of N, where N is the number of aggregated CCs. Cross-CC PF increases the scheduling priority of the single-CC users and improves the fairness (the CDF curves get closer, and the maximum user throughput is closer to the average value). However, it is not completely fair to different kinds of users, because the aim is to maximize the network utility, and not the fairness.

4.4.2 Performance With Finite Buffer Model

For finite buffer traffic model, the average cell throughput equals the offered load as long as the cell is able to accommodate all the arrived users. Otherwise, it

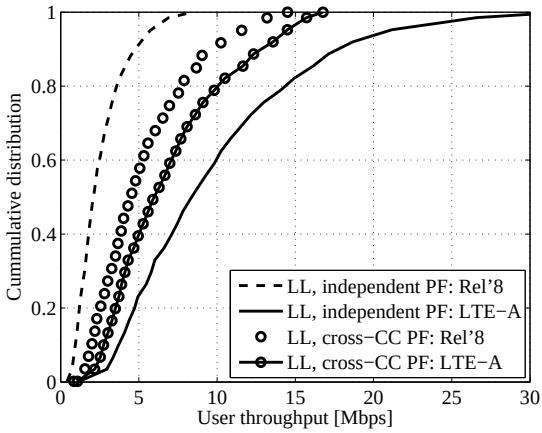


Fig. 4.8: CDF for LTE-Rel'8 and LTE-Advanced users with different PF schedulers. LL balancing is used. 10 full buffer users exist per cell and on average 50% of them are LTE-Advanced.

reaches the capacity limit of 49 Mbps in the considered LTE-Advanced system with 40 MHz bandwidth and 1×2 antenna configuration [108]. This average cell throughput is hardly affected by the different PF algorithms, and hence only the cell edge user throughput is shown. Two cases are considered, the first case with 50% of the users being LTE-Advanced, and different arrival rates. The second case assumes an average load of 20 arrivals per second per cell, but with different ratios of LTE-Advanced users. The results are summarized in Fig. 4.9 and Fig. 4.10, respectively. The performance for MH is also presented.

From Fig. 4.9 it can be seen that the cell edge user throughput decreases with the arrival rate. The reason is that as the arrival rate increases, the total transmission resources are shared among an increasing number of users (see Fig. C.4). Fig. 4.10 shows that when the cell load is fixed, a higher ratio of LTE-Advanced users gives a corresponding improvement in cell edge user throughput. The reason is that with finite buffer transmission, LTE-Advanced users on average finish their data transmissions faster than Rel'8 users; as the ratio of LTE-Advanced users increases, a larger number of users are served in a shorter time, leaving more resources for the remaining users and a better cell edge user throughput is achieved. It is also observed that MH with only Rel'8 users has 1/4 the performance of LTE-Advanced users with independent scheduling. The reason is presented in Section C.3.

Both Fig. 4.9 and Fig. 4.10 confirm that when some users cannot access all the CCs, LL offers a better performance than MH. This is the same as observed with the full buffer traffic model. Secondly, cross-CC PF offers up to 40% gain over independent PF as long as there exists the mixture of the LTE-Advanced and

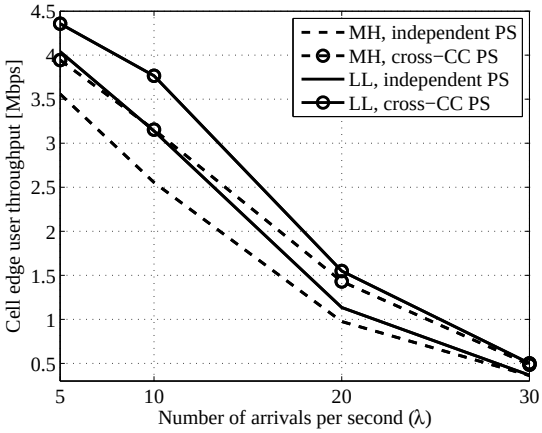


Fig. 4.9: Cell edge user throughput for independent or cross-CC PF with different load balancing methods. Different arrival rates are investigated with 50% LTE-Advanced users.

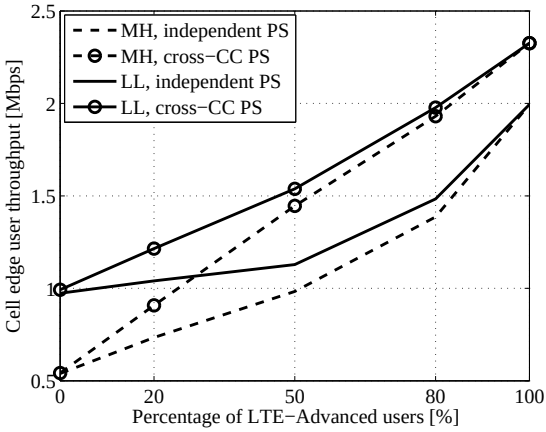


Fig. 4.10: Cell edge user throughput for independent or cross-CC PF with different carrier load balancing methods. On average 20 users arrive per second within each cell, with a variable ratio of LTE-Advanced users.

Rel'8 users. The exact gain depends on the system configurations, traffic model, arrival rate and the LTE-Advanced user ratio.

4.5 Conclusion

In this chapter the problem of how to optimize the resource allocation in a multi-CC system has been studied. The aim is to provide the baseline performance when users are provided with the maximum number of CCs that they can support. Focus is first put on different layer-3 carrier load balancing methods, which assign the CCs to each user. Both theoretical and simulation results are obtained, which show that with low number of users and low percentage of LTE-Advanced users, the load balancing method of LL achieves better performance than the MH balancing. In all situations, the performance of LL is no worse than MH, which makes LL a better solution than MH for the load balancing. This is aligned with similar studies for multi-carrier CDMA systems [53, 55].

After the assignment of CCs, the scheduling of resources in layer-2 is investigated. Using independent PF per CC is found to suffer from low cell edge user throughput. A cross-CC PF algorithm is proposed. With the mixture of both LTE-Advanced and Rel'8 users, cross-CC PF significantly improves the system fairness and cell edge user throughput over independent scheduling. In an LTE-Advanced system, the gain in cell edge user throughput over independent PF is up to 90% with full buffer transmission and 40% with finite buffer transmission, depending primarily on the ratio of LTE-Advanced users. Despite the high gain in cell edge user throughput, it gives no degradation in the average cell throughput and maximizes the log-measure network utility, and is therefore an attractive candidate for the layer-2 PS.

The results presented in this chapter also indicate that at low to medium cell load, the 'carrier aggregation' mode when each user can access all CCs gives better performance than the '$N\times$independent carrier' mode, where only one CC is accessible per user. However, they achieve the same performance when the cell load is high. Therefore, it is preferable to assign less CCs to the LTE-Advanced users at high load. This reduces the uplink feedback overhead and user power consumption. Detailed analysis on feedback reduction will be provided in the next chapter.

Chapter 5

Downlink Performance of Multi-CC Systems With Reduced Feedback

5.1 Introduction

When channel aware packet scheduling and link adaptation are used, the system performance can be significantly improved as compared to the case without the channel knowledge [24–26]. A feedback link and transmission resources are required to provide the transmitter with the Channel Quality Information (CQI), and the Acknowledgement (ACK)/Non-acknowledgement (NACK), which tells whether or not the reception is correct [27]. For codebook based MIMO transmission, the base station also needs the information on which precoding matrix to use (the Precoding Matrix Indicator (PMI)), and the number of supported data streams (the Rank Indication (RI)) [27].

In multi-Component Carrier (CC) systems, if full CQI and ACK/NACK are reported on each CC, they may contribute to a large amount of transmission overhead. A high overhead is disadvantageous in several aspects: first, it consumes uplink resources that could otherwise be used for data transmission; secondly, such control data requires high transmit power which may be infeasible for the cell edge users; last but not least, it increases the operation complexity in both obtaining the data and dealing with them. In order to reduce the resources occupied by the

feedback information, many feedback reduction techniques have been developed. References [59–61] reduce the CQI overhead by reporting a quantized discrete value instead of the full channel state. A compression technique based on a discrete cosine transform is presented in [62]. In OFDMA systems, because a user is most likely to be scheduled on the resource blocks with good channel quality, a best-M scheme that reports only the best resource blocks gives good performance [62, 63]. It can be implemented in several ways, e.g. individual best-M, which feeds back the CQI value for each selected block, or average best-M, which feeds back the average CQI for the selected ones. The latter has been adopted by the 3GPP for LTE [27, 64]. The feedback can also be reduced by increasing the feedback interval when mobility is low [65]. For the ACK/NACK signaling, a technique called ACK/NACK bundling has been developed to reduce this overhead [66]. It combines the reception indications of multiple packets into one ACK/NACK [27], and the performance in an LTE TDD system has been evaluated in [67].

This chapter is devoted to reducing the overhead in two aspects: the CQI and the ACK/NACK signaling. Consequently, it contains two main sections. Section 5.2 focuses on the CQI reduction. It starts with investigating the appropriate CQI reporting methods within each CC. Afterwards, several cross-CC CQI reduction techniques are developed, and their performance is compared with each other. A load adaptive scheme is proposed at the end of this section. In Section 5.3, the ACK/NACK bundling technique, which was originally proposed for TDD systems, is applied to combine the ACK/NACKs across the CCs. A simple analysis of the uplink range limitation for different feedback techniques is provided in Section 5.4. Finally, the main findings in this chapter are concluded in Section 5.5.

5.2 CQI Reduction Techniques and Their Performance

5.2.1 Introduction to CQI Reduction Techniques

CQI is a quantized SINR value (in decibel) which is essential for a system to exploit the channel variation using packet scheduling and link adaptation. There are 50 Physical Resource Block (PRB)s in a 10 MHz CC [22]. When best-M is used, three neighboring PRBs are grouped and associated with one CQI value [27]. This group size is used here for all the reporting methods. Based on [109], 4 bits are used for the quantization in this study. Together with a quantization step of 1.6 dB, a dynamic range of 25 dB is obtained, which is sufficient for link adaptation (see Fig. A.7). This study also assumes a feedback interval of 5 ms. The existing techniques on CQI reduction are for the single-CC system, hence named as intra-CC CQI reduction in this thesis. In a multi-CC system, it is possible to use

different reporting methods at different CCs, leading to inter-CC CQI reduction. The different CQI reduction techniques considered in this study are described below.

5.2.1.1 Intra-CC CQI Reduction

Intra-CC CQI reduction is achieved by using a proper reporting method within each CC, instead of feeding back the exact channel state for each CQI group. It has been extensively studied in the literature, as discussed in the introduction. In this study, because the SINR quantization step and the feedback interval have been fixed, only the best-M method is selected from the available options. As a special case, reporting only the wideband CQI is also investigated, which corresponds to best-0 report. A short introduction to the selected techniques is provided here.

Full CQI: One CQI value is sent back for each CQI group. This is the reference case with the highest amount of overhead. It requires $N_{CQI} \times Q_{CQI}$ bits for the feedback per CC, where N_{CQI} is the total number of CQI groups in the CC; $Q_{CQI} = 4$ is the quantization step.

Best-M: The average best-M is selected. It works as follows [27]:

The user first estimates the channel quality and then calculates the average CQI value of the best M_{CQI} groups. This average CQI is encoded differentially using $Q_\Delta = 2$ bits relative to the wideband CQI W: $\Delta_{bestM} = CQI_{bestM} - W$. The feedback contains the offset value, the wideband CQI, and the position of the selected CQI groups. The overhead is therefore:

$$Q_{bestM} = Q_\Delta + Q_{CQI} + \left\lceil \log_2 \binom{N_{CQI}}{M_{CQI}} \right\rceil \tag{5.1}$$

In (5.1), $\lceil \cdot \rceil$ is the ceil function that rounds a value to the nearest larger integer. Because each user selects the CQI groups based on its own channel condition, there is a risk that some groups are not associated with any CQI report. Several methods have been developed to utilize these unclaimed groups [65,109,110], and a simple one is used here. It assumes the unclaimed CQIs have the same CQI value, which differentiates from the average CQI by an offset Δ_{worstM}. In this way, the average CQI of the selected and unclaimed CQI groups equals the wideband CQI, and hence:

$$N_{CQI}W = M_{CQI}(W + \Delta_{bestM}) + (N_{CQI} - M_{CQI})(W - \Delta_{worstM})$$

W appears in both sides of the equation and cancels each other. Therefore, Δ_{worstM} only depends on the value of M_{CQI} and the offset of the good CQI groups: $\Delta_{worstM} = M_{CQI}\Delta_{bestM}/(N_{CQI} - M_{CQI})$.

Wideband CQI: Only the wideband CQI is report, which does not contain the frequency domain diversity. The number of required bits is Q_{CQI}.

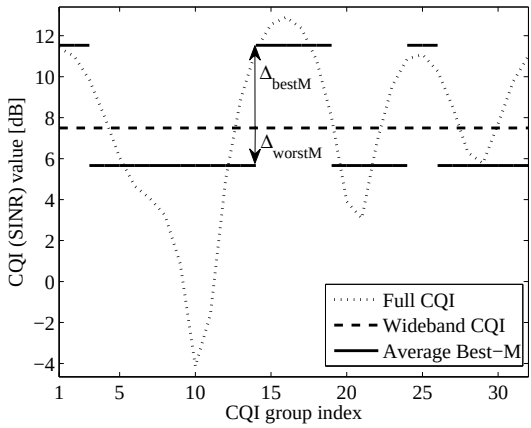


Fig. 5.1: Example of full CQI, wideband CQI and average best-M feedback.

Fig. 5.1 gives an example of the different CQI reporting methods. It is worth mentioning that the performance of average best-M can be improved by adapting the value of M_{CQI} according to the cell load. Detailed investigation of load adaptive average best-M is presented in Section D.2. Based on the finding, $M_{CQI} = 5$ achieves very close performance to the adaptive case in the considered LTE-Advanced system, with an exception that $M_{CQI} = 0$ (wideband CQI) is enough to achieve the best performance when there is only a single user per cell. Therefore, a fixed M_{CQI} of 5 is used in this chapter for best-M, with the possibility to switch to wideband CQI report at low load.

5.2.1.2 Inter-CC CQI Reduction

In a multi-CC system, different CQI reporting methods could be used for the LTE-Advanced users across the CCs, offering more flexibility to the overhead reduction than in a single-CC system. It will be seen later that wideband CQI and best-M offer attractive performance than full CQI report. Therefore, they are considered as candidates for overhead reduction in each CC. Furthermore, the possibility of blocking a user on certain CCs is also considered. The user will not be scheduled on the blocked CCs and need not to report the CQI. The combination of these methods in different CCs leads to the following schemes:

Best-M in all CCs: It repeats the best-M procedure across the aggregated CCs and therefore has no cross-CC reduction. Assuming equal bandwidth for

the CCs, the feedback overhead is $N \times Q_{bestM}$, where N is the number of aggregated CCs and Q_{bestM} is the overhead for one CC, calculated by (5.1).

Best and Wideband CQI (B&W): It dynamically selects the best N_{BM} CCs and uses best-M for these CCs. N_{BM} is an integer between 0 and N. The remaining CCs are indicated with the wideband CQI. B&W is simple to implement because the selection is performed by the user, no coordination with the others is needed. $N_{BM} = N$ is the same as having best-M in all CCs, and $N_{BM} = 0$ corresponds to wideband CQI report per CC. Among various selection metrics, the wideband CQI is found to be a good one for selecting the CCs. However, because of the averaging effect, if the bandwidth is so large that different CCs have similar wideband CQI, another selection metric should be used.

The feedback of B&W constitutes of three parts: the bits indicating the selected CCs, the best-M feedback for these CCs, and the wideband CQI for the remaining CCs. The feedback overhead can be expressed as:

$$Q_{B\&W} = \left\lceil \log_2 \binom{N}{N_{BM}} \right\rceil + N_{BM} Q_{bestM} + (N - N_{BM}) Q_{CQI} \tag{5.2}$$

Best-CC (BCC): The BCC algorithm is a simplification of B&W. It allows a user to operate only on the best CCs (with highest wideband CQI), and the remaining CCs are blocked. The motivations for this simplification are:

1. To further reduce feedback overhead.
2. The wideband CQI appears at the transmitter as a degenerated version of the frequency selective CQIs. It may be beneficial to schedule resources to a user with slightly lower, but more accurate CQI, than to a user with an inaccurate wideband CQI. If this is the case, the additional feedback of wideband CQI on the worst CCs brings no gain on system performance.

BCC requires only the first two components of (5.2), i.e.:

$$Q_{BCC} = \left\lceil \log_2 \binom{N}{N_{BM}} \right\rceil + N_{BM} Q_{bestM} \tag{5.3}$$

Balanced CC Load (BCL): The BCL algorithm is based on the Least Load (LL) balancing, which is discussed in Chapter 4. While LL is applied to LTE-Rel'8 users, and selects only one CC per user, BCL selects multiple (N_{BM}) CCs for each LTE-Advanced user and tries to balance the load (number of users) with respect to the CC bandwidth. For the purpose of load balancing, BCL should be carried out at the base station, which knows the selected CCs for each user. This is different from BCC, which operates at the user side. BCL is simpler than BCC because the user needs not to monitor the channel

quality on all CCs. Because the CC assignment is fixed for each user, only the CQI reports for the selected CCs are needed, and hence:

$$Q_{BCL} = N_{BM} Q_{bestM} \tag{5.4}$$

It is worth noting that BCC and BCL are in fact CC selection methods, which achieve CQI reduction by selecting a subset of the available CCs for each user.

5.2.1.3 Overhead Analysis of CQI Reduction

According to the system setup in Table A.3, 4 CCs are aggregated, each has 10 MHz bandwidth, corresponds to 50 PRBs. With a group size of 3, there are $\lceil 50/3 \rceil = 17$ CQI groups per CC. Using the equations in Section 5.2.1.1, the CQI overhead per CC for the different reporting methods can be calculated, which is 68 bits for full CQI; 19 bits for best-M; and only 4 bits for wideband CQI.

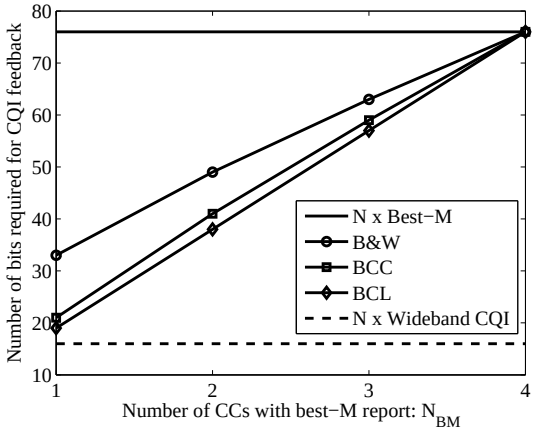


Fig. 5.2: Number of CQI bits versus the number of CCs using best-M report.

For the different inter-CC CQI reduction techniques, their CQI overheads can be calculated by using (5.2) to (5.4), and are plotted in Fig. 5.2. This figure shows that the overheads for B&W, BCC and BCL increase roughly linearly with the number of CCs that are reported with best-M. $N_{BM} = 4$ means best-M is used in all CCs; it has the maximum overhead of 76 bits and is denoted by 'N×Best-M' in the figure. B&W with $N_{BM} = 0$ is the same as having wideband CQI report per CC, which requires only 16 bits and is denoted by 'N×Wideband CQI'.

5.2.2 Performance With Intra-CC CQI Reduction

Having described the different CQI reduction techniques, this subsection and the following one analyze their performance. The analysis is carried out based on system level simulations, using the same assumptions as listed in Table A.3. The only difference is that, in order to reduce the complexity of the frequency domain scheduling and link adaptation, maximum 10 users are allowed to pass through the time domain scheduler. Since the overhead reduction is less critical for LTE-Rel'8 users that operate on a single CC, only the LTE-Advanced users are considered. The results are shown only for the finite buffer model. The performance with the full buffer model has similar trends as the finite buffer model and can be found in Appendix E. Various arrival rates have been considered, ranging from 5 to 30 arrivals per cell per second. Together with a buffer size of 2 Mbits, the simulation covers an offered load from 10 to 60 Mbps.

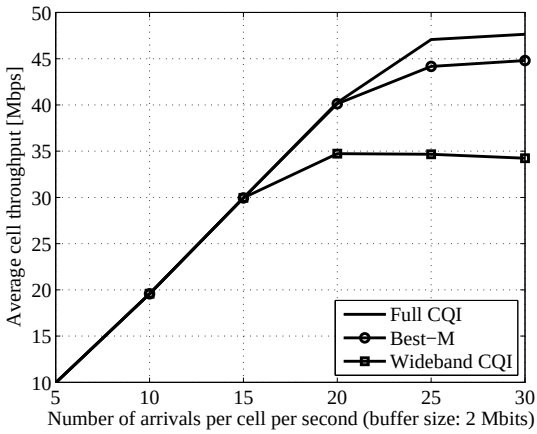


Fig. 5.3: Average cell throughput with intra-CC CQI reduction.

The average cell throughput with intra-CC CQI reduction is plotted in Fig. 5.3. As can be seen from this figure, for an arrival rate no larger than 15 (offered load $\leq$ 30 Mbps), all investigated CQI reporting methods are able to serve the offered load. Therefore, they have the same average cell throughput, which equals the offered load. As the arrival rate further increases, wideband CQI cannot offer sufficient capacity, and the average cell throughput saturates at around 35 Mbps; with best-M report, the system can serve up to 22 arrivals per second, and the cell capacity is 45 Mbps; full CQI report gives the highest capacity of 48 Mbps. Note that the capacity achieved by full CQI is 2% lower than in [108](49 Mbps). This is due to the limit of maximum 10 users for frequency domain scheduling. The loss can also be estimated by using (B.21), as: $1-\left(0.11ln(10)+1.10\right)/1.38=2\%$, where 1.38 is the maximum multi-user gain when more than 13 users are available

for frequency domain scheduling.

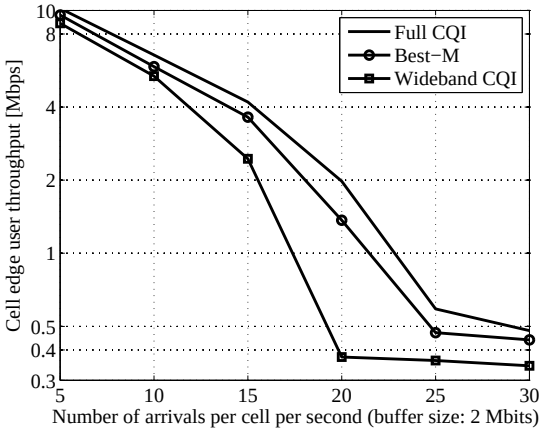


Fig. 5.4: Cell edge user throughput with intra-CC CQI reduction.

For the cell edge user throughput plotted in Fig. 5.4, the following observations can be made:

1. At low arrival rates, different feedback methods have similar performance. The reason is that when the arrival rate is very low, the active user is most likely to be scheduled on the whole bandwidth. Only the wideband CQI value is used, even if the feedback contains more detailed information.

2. As the arrival rate increases, the packet scheduler can benefit from frequency domain diversity if full CQI or best-M is used. Therefore the performance is better than with only wideband CQI. Best-M achieves 70∼100% the performance of full CQI report. With an arrival rate of 20, the offered load is beyond the capacity of wideband CQI, but below that of full CQI, using wideband CQI gives only 19% the cell edge user throughput of full CQI report.

Taking both the downlink performance and the uplink overhead into consideration, best-M offers a better trade-off between performance and overhead than full CQI. It reduces the overhead by 72% while retaining good downlink performance. Wideband CQI reduces the feedback by 94%, it is recommended when the cell is lightly loaded.

5.2.3 Performance With Inter-CC CQI Reduction

5.2.3.1 Performance Comparison of Different Techniques

After studying CQI reduction within a CC, the performance of inter-CC CQI reduction is analyzed in a system with 4 CCs. For the average cell throughput, different CQI reduction techniques offer the same performance as long as the offered load lies below the cell capacity. Similar as shown in Fig. 5.3, the average cell throughput is bounded between 35 Mbps ('N×Wideband CQI') and around 50 Mbps. Therefore, it is not plotted here. One thing worth mentioning is that, the highest cell capacity is achieved by assigning the best CC to each user (BCC with $N_{BM} = 1$) instead of having frequency selective CQIs on all CCs (N×Full CQI or Best-M). This is due to the fact that only 10 users are allowed for the frequency domain scheduling, and the time domain scheduling is channel blind. Using only the best CC for each user exploits the frequency domain diversity across the CCs, and hence improves the average channel quality within each CC.

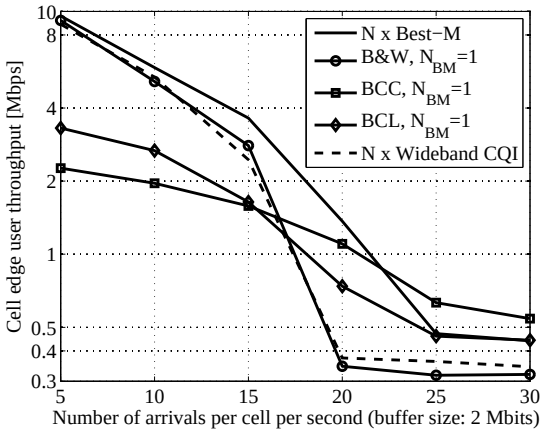


Fig. 5.5: Cell edge user throughput with inter-CC CQI reduction, $N_{BM} = 1$.

The cell edge user throughput is affected not only by the offered load, but also by the way the resources are utilized, and therefore it is sensitive to the feedback reduction techniques in all load conditions. In Fig. 5.5, the cell edge user throughput of different feedback reduction techniques for $N_{BM} = 1$ is shown, with the corresponding average number of active users plotted in Fig. 5.6. Fig. 5.5 indicates that at low load the preferable technique is using only wideband CQI, whereas at high load either BCC or BCL can be used. BCC exploits the multi-CC diversity and offers the highest cell edge user throughput at high load. However, its performance is poorer than BCL at low load. This is because BCC selects the CCs based on the channel quality but not the load condition. As a result, more than one users

may select the same CC even if there are empty CCs. BCL ensures maximum one user per CC as long as $KN_{BM} \leq N$, where K is the number of users in a cell. Therefore it is better than BCC at low load.

It is also observed from Fig. 5.5 that B&W has similar performance as $N\times$Wideband CQI at low or medium load, whereas at high load the performance of B&W is worse than using wideband CQI in all CCs. The reason for the poor performance of B&W at high load is: in wireless transmissions, the deep fades can be suppressed by either coding across a very wide bandwidth or using only the good spectrum. However, if the mixture of both wideband and frequency selective CQIs exists (which is often the case when B&W is used), the scheduler tends to assign each user a narrower bandwidth. For the users with only wideband CQI, they do no have enough resistance to the deep fades. If the frequency domain diversity offered by B&W cannot compensate for the loss due to deep fades, poor throughput performance will be experienced. Overall, $N\times$Wideband CQI is more preferable than B&W, as it gives similar or better performance than B&W with smaller overhead.

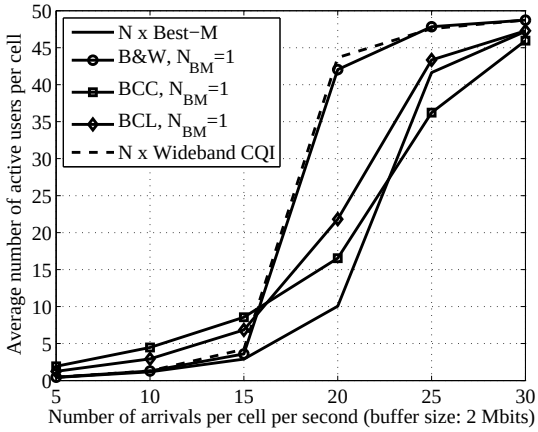


Fig. 5.6: Average number of active users with inter-CC CQI reduction, $N_{BM} = 1$.

From the average number of active users shown in Fig. 5.6, it can clearly be seen that as the arrival rate increases, more users will remain active in the system. When the arrival rate is so high that the system cannot support the load, the number of active users will be limited by admission control, at maximum 50.

5.2.3.2 Performance With a High Value of N_{BM}

As discussed in Section 5.2.1.2, it is in general possible to improve the downlink performance by allowing more CCs to be reported with best-M ($N_{BM} > 1$).

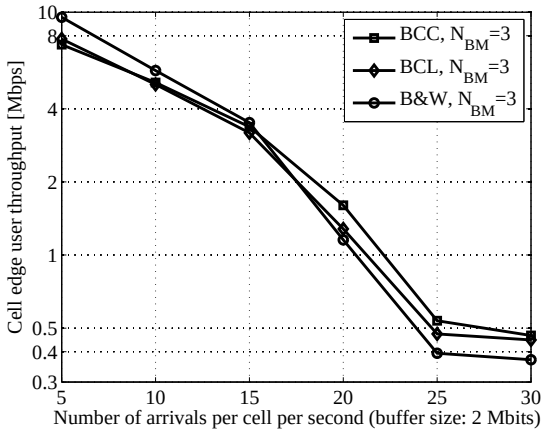


Fig. 5.7: Cell edge user throughput with inter-CC CQI reduction, $N_{BM} = 3$.

Meanwhile, the feedback overhead will rise. Fig. 5.7 summarizes the cell edge user throughput when each user has best-M report on three CCs ($N_{BM} = 3$). Comparing Fig. 5.7 with Fig. 5.5 it can be seen that:

1. At low load, the performance of BCC and BCL is significantly improved by increasing the value of N_{BM}, because each user can simultaneously access a much wider bandwidth and use more transmission resources per TTI.

2. At high load, BCC and BCL cannot benefit from a high N_{BM}. This is due to the fact that when many users exist, even if each user is assigned only one CC, there will be enough number of users (≥ 10) on each CC to be exploited by the frequency domain packet scheduler.

3. B&W has a different behavior from BCC and BCL. At very low load, there is hardly any gain by increasing the value of N_{BM} because wideband CQI is enough. At high load, a high N_{BM} increases the possibility that a user is scheduled based on frequency selective CQIs. So the performance is increased. However, even at high load, the performance of B&W is worse than the other two schemes, despite the high overhead.

From these observations, it is suggested to use $N\times$Wideband CQI at very low load; as load increases, either BCC or BCL can be used, and the value of N_{BM} should decrease with the arrival rate.

5.2.4 Overall Performance and Recommendations for CQI Reduction

According to the findings before, by decreasing the value of N_{BM} with the cell load (arrival rate), and choosing between wideband CQI and best-M on each CC, the CQI overhead can be significantly reduced at reasonable loss of the downlink throughput. Fig. 5.8 summarizes the relative performance and CQI reduction as compared to the case with full CQI on each CC. The selection of the CQI reduction technique at each load condition is based on the simulation results. As can be seen from this figure, the overhead is reduced by 79∼94%, while achieving higher than 75% the performance in most cases. There is an exception when the arrival rate approaches the network capacity. In this situation reducing the CQI leads to much higher number of active users than with full CQI report, and hence the cell edge user throughput is poorer.

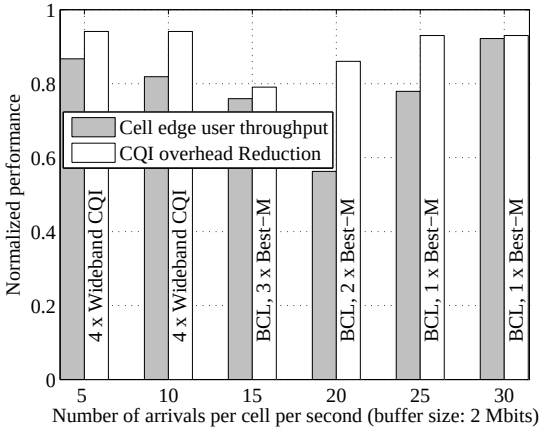


Fig. 5.8: Throughput and overhead reduction of load adaptive CQI reduction.

Fig. 5.8 demonstrates the usefulness of the load adaptive CQI reduction. However, it requires numerous simulations to select the reduction technique at each load condition. If the network configuration is changed, e.g., a different carrier aggregation pattern is used or the users have a different buffer size, the suitable CQI reporting method at a specific load condition may be different from the one used here. In order to automatically select the proper CQI reduction technique in various situations, a load adaptive scheme is proposed.

A flowchart of the load adaptive scheme is depicted in Fig. 5.9. As shown, the users are separated according to their terminal categories: the LTE-Advanced users are assigned to N_{BM} CCs, using BCL or BCC; the LTE-Rel'8 users can access only

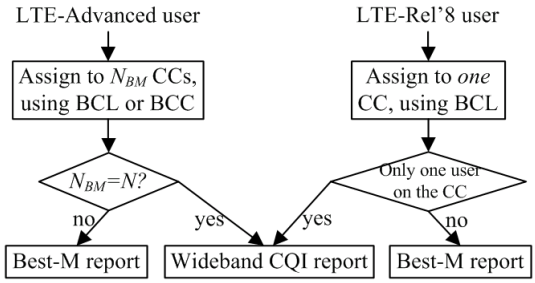


Fig. 5.9: Load adaptive CQI reduction in multi-CC systems.

one CC, they do not support the fast switching between CCs and therefore are statically assigned to the CC that has the lowest load (BCL, $N_{BM} = 1$). The number of CCs assigned to an LTE-Advanced user is:

$$N_{BM} = min\left\{max\left\{1, NT - \sum_{i=1}^{K} N_{BM,k}\right\}, N\right\} \tag{5.5}$$

In (5.5), N is the number of aggregated CCs, T is the number of users to achieve enough multi-user diversity, K is the number of active users, and $N_{BM,k}$ is the number of CCs assigned to the existing user k. The goal of (5.5) is to ensure T users per CC whenever possible. According to Fig. B.6, a reasonable value of T in a macro-cell scenario is between 6 and 10.

In general, N_{BM} decreases with the number of active users K, meaning that at high load, users can be assigned with less CCs without losing system performance. The assignment of CCs can also be done in an iterative manner, which loops over the CCs and selects the one that has the lowest load within each iteration. After selecting the CC(s), a proper CQI reporting method will be decided. If $N_{BM} = N$, this indicates that the cell is lightly loaded, and hence only wideband CQI is needed; otherwise, best-M should be used on the selected CCs, and the remaining ones are blocked. For the LTE-Rel'8 user, it selects wideband CQI if there is only one user on the CC, or best-M when multiple users exist.

5.3 ACK/NACK Bundling Techniques and Their Performance

5.3.1 Introduction to ACK/NACK Bundling

ACK/NACK is an on/off status that can be sufficiently indicated with one binary

bit. However, it requires high reliability of the feedback channel and uses very strong coding. Therefore, it may also contributes to a large amount of uplink overhead.

The signaling overhead of ACK/NACKs can be reduced by combining the multiple ACK/NACKs for each user: an ACK is reported if all the transmission blocks are correctly received; otherwise, a NACK is sent back and retransmission will take place for all the blocks. This is known as ACK/NACK bundling [27,66]. For one user, the potential of bundling comes from 1) the time domain due to the continuous transmission/reception, 2) the CC domain, because a user is scheduled on multiple CCs, 3) the antenna domain, when more than one data streams are multiplexed to each user. The benefit of time domain bundling has been investigated in [67]. Therefore, only CC and antenna domain bundling is considered in this study, leading to the following schemes:

No bundling: This is the reference which sends full ACK/NACKs in the uplink.

Intra-CC bundling: This scheme is applicable for MIMO systems, where each user indicates its reception status on one CC with a single ACK/NACK, irrespective of the transmission rank:

$$\Psi_n = \prod_{1 \leq m \leq M} \Psi_{n,m}, \quad \forall\ 1 \leq n \leq N$$

M is the highest MIMO rank (maximum number of data streams); n is the CC index. $\Psi_{n,m}$ is the reception status indicator at the n^{th} CC for the m^{th} MIMO stream. It has the value of 0 if reception is in error, otherwise, $\Psi_{n,m} = 1$. Ψ_n is the product ('AND') of the receiver side indicators across the MIMO streams. After the bundling, up to N ACK/NACKs are reported from each user.

Inter-CC bundling: This scheme uses the 'AND' operation over the same MIMO stream across multiple CCs, and requires maximum M ACK/NACKs, according to the supported MIMO rank:

$$\Psi_m = \prod_{1 \leq n \leq N} \Psi_{n,m}, \quad \forall\ 1 \leq m \leq M$$

Full bundling: This is a combination of intra- and inter-CC bundling, which uses one ACK/NACK for all CCs and all possible MIMO streams:

$$\Psi = \prod_{\substack{1 \leq n \leq N \\ 1 \leq m \leq M}} \Psi_{n,m}$$

Partial bundling: Partial bundling is a generalized scheme of the full bundling, where multiple ACK/NACKs are used. The ACK/NACKs within a CC

are always combined using intra-CC bundling (later it is shown that intra-CC bundling gives nearly no performance loss). An integer number of N' ACK/NACKs are used for the signaling, $1 \leq N' \leq N$. $N' = 1$ is the same as full bundling, and $N' = N$ leads to intra-CC bundling.

The different ACK/NACK bundling techniques are illustrated in Fig. 5.10.

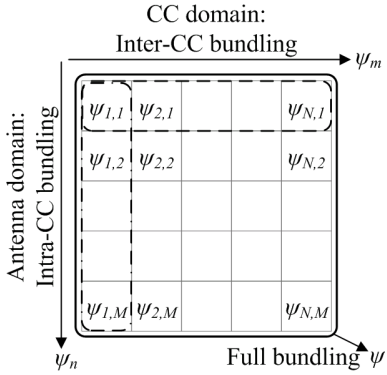


Fig. 5.10: CC domain and antenna domain ACK/NACK bundling.

The status of the packet reception determines the BLER. At the receiver side, BLER can be calculated based on the reception of the individual data stream, and is named $BLER_{rx}$. However, the transmitter only knows the status of the bundled receptions, and hence has a different BLER, denoted by $BLER_{tx}$. In case of fully uncorrelated receptions and the same $BLER_{rx}$ for the different data streams, the relationship between the transmitter side and the receiver side BLER can be expressed as:

$$BLER_{tx} = 1 - (1 - BLER_{rx})^{N_S/N_A} \tag{5.6}$$

where $N_S \leq M \times N$ is the number of simultaneously received data streams of a user; N_A is the number of ACK/NACKs used for uplink feedback. $N_A \leq N_S$ because the number of ACK/NACKs reported to the transmitter should be no larger than the number of simultaneous transmissions. Equation (5.6) shows that $BLER_{tx} \geq BLER_{rx}$, and hence there will be unnecessary retransmissions for some correctly decoded blocks. Moreover, in order to maintain the BLER target at the transmitter side, the lower value of $BLER_{rx}$ leads to a much higher outer-loop link adaptation offset (in order to satisfy the higher SINR requirement at the receiver side) than the case without bundling. As a consequence, lower modulation and coding schemes are favored and the spectral efficiency is penalized. Therefore, the trade-off between overhead reduction and performance degradation exists, which is analyzed in the rest of the chapter.

Both Single Input Multiple Output (SIMO) and MIMO are used for the performance evaluation. The reason for considering MIMO is that the intra-CC bundling is only applicable when multiple data streams are simultaneously sent to a user on one CC. The results are generated using both full and finite buffer traffic models.

5.3.2 Performance of ACK/NACK Bundling in SIMO

In a SIMO system, intra-CC bundling is inapplicable, so comparison is only made between the performance of no bundling, partial bundling and full bundling. Best-M on all CCs is assumed here, i.e., no cross-CC CQI reduction.

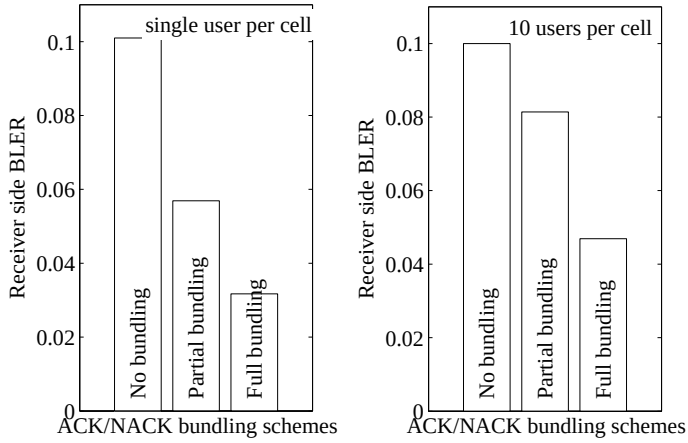


Fig. 5.11: Receiver side BLER with ACK/NACK bundling using full buffer model.

The receiver side BLER from simulation is plotted in Fig. 5.11 for both the single user case and the multi-user case with 10 user per cell. For the single user case, the user will be scheduled on all CCs, i.e., $N_S = N = 4$. In order to achieve the 10% BLER target at the transmitter side, the calculated receiver side BLER using (5.6) is 5.1% when $N_A = 2$, and 2.6% when $N_A = 1$. When multiple users exist, a user may not be simultaneously scheduled on all CCs. As a result, the bundling is performed across a smaller number of ACK/NACKs than the single user case and the receiver side BLER should be higher. Fig. 5.11 matches well the discussion here, indicating the implementation of ACK/NACK bundling is correct.

The downlink throughput is summarized in Fig. 5.12, with full buffer traffic model and number of users per cell varying from 1 to 10. As can be seen from this figure,

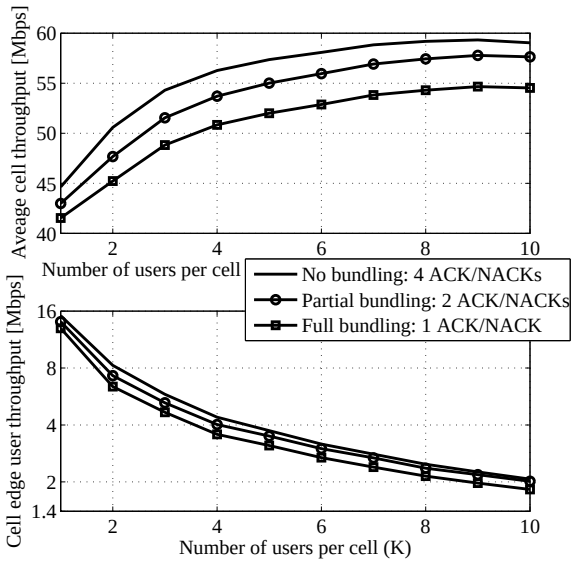


Fig. 5.12: Downlink throughput with ACK/NACK bundling and full buffer model.

the performance decreases when the number of ACK/NACKs is reduced. Also, the loss of ACK/NACK bundling gets smaller as the number of user increases. E.g., when the number of users per cell is small (less than 3), partial bundling with 2 ACK/NACKs per users gives around 5% loss in average cell throughput and 10% loss in cell edge user throughput; these losses decrease to only 2.5% in both average cell throughput and cell edge user throughput with 10 users per cell. When only one ACK/NACK is used and the number of users is small (less than 3), the losses are about 10% in average cell throughput and 20% in cell edge user throughput; with 10 users, the losses are 7.5% and 12%, respectively.

When the finite buffer traffic model is used, similar behavior as with the full buffer model is experienced, as shown in Fig. 5.13. When two ACK/NACKs are used for the feedback, the performance is 10∼0% lower than no bundling. In the case of full bundling, the loss is 22% at low load and 7% at high load.

5.3.3 Performance of ACK/NACK Bundling in MIMO

A 2×2 antenna configuration is taken for the evaluation of the ACK/NACK bundling in MIMO systems. When the channel quality is good, up to two data streams can be multiplexed to each user. Otherwise, rank-1 transmission is employed to exploit the MIMO diversity. Both rank-1 and rank-2 transmissions are

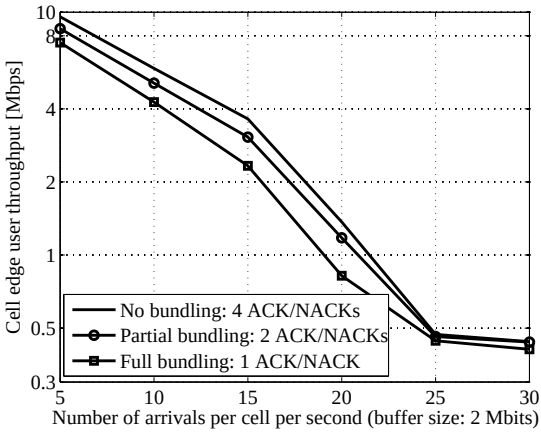


Fig. 5.13: Cell edge user throughput with ACK/NACK bundling. Finite buffer model is used with different arrival rates.

realized in the closed-loop manner with uplink feedback. One can refer to [111] for detailed information about the MIMO technologies in LTE.

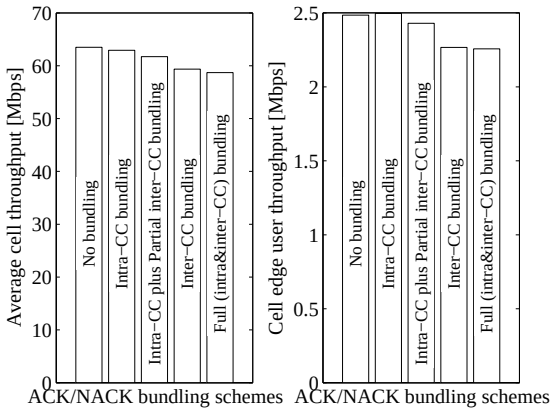


Fig. 5.14: Downlink throughput of MIMO system with ACK/NACK bundling. Full buffer model is used with 10 users per cell.

Simulating the MIMO channels and the precoding process is really memory consuming, and it takes long simulation time. Therefore, only the 10 user case is simulated, and the result is plotted in Fig. 5.14. As shown, intra-CC bundling gives nearly no loss in the downlink performance. The reasons are:

1. Only a small portion of the users can support multi-stream transmission in the given macro-cell scenario. According to the simulation results, only around 10% of the transmissions are rank-2. This is the most important reason for the very small performance loss of intra-CC bundling.

2. The receptions for the multiple data streams are not fully uncorrelated, i.e., if one data stream is received in error, there is a high probability that the other stream is also in error. To demonstrate this, the simulated probabilities of the different reception status of the rank-2 transmissions are plotted in Fig. 5.15. It shows the probabilities of the three possible events: "none of the streams is in error" (P_0), "only one of the streams is in error" (P_1), and "both streams are in error" (P_2). When one stream is received in error, the probability of the other being in error is: $P_2/(P_1 + P_2) = 17\%$. This is higher than the overall reception error rate (per data stream) of $P_e = P_1/2 + P_2 = 12\%$, indicating a small correlation between the multiplexed streams. Besides, if the two streams are fully uncorrelated, the probability of having both streams in error would be $P_e^2 = 1.5\%$. This is lower than the simulated probability of 3.5%. Bundling the correlated ACK/NACKs causes less unnecessary retransmissions than uncorrelated receptions, and hence the loss of intra-CC bundling as compared to no bundling is very small.

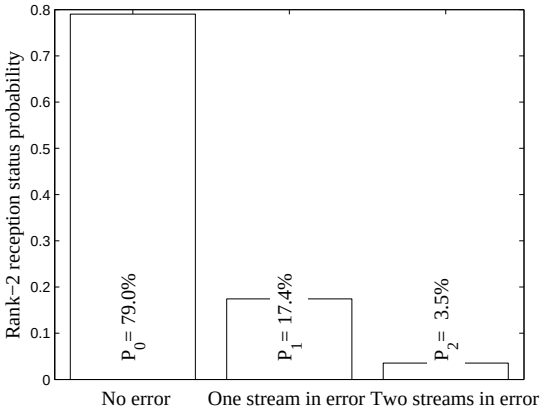


Fig. 5.15: Reception error probabilities for rank-2 transmissions. Full buffer model is used with 10 users per cell.

As the number of ACK/NACKs decreases further, the downlink performance suffers from a degradation of maximum 7.6% in average cell throughput and 9.2% in cell edge user throughput (with full bundling). This degradation is accompanied by an ACK/NACK overhead reduction of $1 - 1/(NM) = 87.5\%$.

5.4 Range Limitation of Different Feedback Methods

In this section, the range limitation of different feedback methods is analyzed. The different feedback methods are obtained by using overhead reduction of both CQI and ACK/NACK. Only the users with poor signal quality are considered, because these users are most likely to be uplink power limited, and will benefit most from overhead reduction. Furthermore, no uplink data transmission is considered. I.e., all the uplink transmissions are for the feedback. This is usually the case when the user is downloading a file or in video streaming mode.

When SINR is very low, it is known from the Shannon's law that the mapping from SINR to throughput is almost linear. Therefore, the amount of uplink overhead is proportional to the SINR requirement. Assuming the same interference level across the cell coverage, the SINR is proportional to the received signal strength, and in turn, the path gain to the serving base station. In equation:

$$Overhead \propto SINR \propto d^{-3.76} \tag{5.7}$$

In (5.7), $\propto$ means 'proportional to', and the exponential coefficient of -3.76 comes from the path loss expression of (A.6).

According to [112], the ACK/NACK bits are repeated 4 times in the uplink to increase reliability. Therefore, when calculating the overall overhead, the contribution from ACK/NACK should be weighted 4 times over CQI. The following cases have been considered:

Case 1: Single-CC transmission with average best-M and one ACK/NACK: the overhead is: $Q_{bestM} + 4 = 23$.

Case 2: Average best-M on all CCs with one ACK/NACK, the overhead is: $NQ_{bestM} + 4 = 80$.

Case 3: Average best-M and full ACK/NACKs are reported on all CCs, the overhead is: $N(Q_{bestM} + 4) = 92$.

Case 4: Full CQI and ACK/NACKs are reported on all CCs, the overhead is: $N(Q_{Full} + 4) = 288$.

The range limitation of the four feedback methods is calculated by using (5.7), and is plotted in Fig. 5.16 (normalized to case 1). As can be seen from this figure, single-CC transmission can be applied in a much wider range than the other cases, due to the considerably lower feedback overhead. Next is case 2 (average best-M on all CCs but only one ACK/NACK), which can be applied within 72% of

the range of case 1. Case 3 uses four ACK/NACKs and further decreases the applicable range by 3%. The smallest applicable range is supported when full CQI and ACK/NACK are to be reported, which is only 51% as compared to single-CC with average best-M. In general, Fig. 5.16 indicates that the amount of overhead a user can support decreases with the distance to the serving base station. It is therefore necessary to take the range limitation into consideration when selecting the feedback method. For example, a user that supports the uplink feedback of only one CC should not be configured to operate on multiple CCs in the downlink.

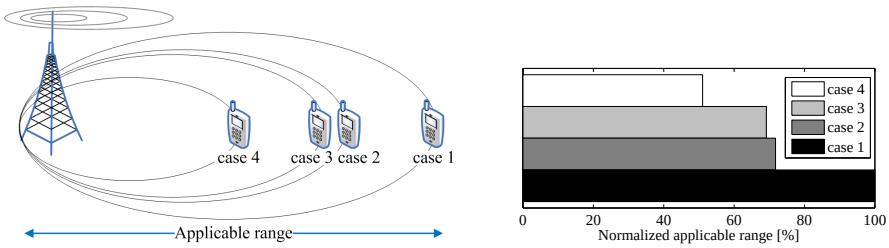


Fig. 5.16: Range limitation of different feedback methods.

Although ACK/NACK bundling gives much smaller overhead reduction compared to CQI reduction methods, it is worth noting that the ACK/NACK is usually transmitted in the uplink control channel, whereas the frequency selective CQI is sent via the uplink data channel. The control channel capacity is in general lower than the data channel, therefore it may still be worth the effort to reduce ACK/NACK by bundling.

5.5 Conclusion

In this chapter the effect of reducing uplink feedback on the downlink performance has been analyzed. The feedback reduction has been performed for two kinds of uplink information: the CQI, which is used for packet scheduling and link adaptation, and the ACK/NACK of the packet reception.

The CQI reduction is performed both within each CC and across the CCs. For the intra-CC reduction, two known methods of average best-M and wideband CQI report have been investigated with full CQI feedback as the reference. It is found that wideband CQI is efficient when there are few users per cell. Otherwise, best-M can be used to offer the frequency domain diversity. For the inter-CC reduction, several techniques have been developed, but there is no unique technique to suit

the transmission in all load conditions. As a result, a load adaptive CQI reduction scheme is proposed. It assigns a smaller number of CCs to the LTE-Advanced users when the load is high, and vice versa. The proposed scheme also includes the selection between wideband CQI and average best-M. As compared to full CQI report, it reduces the CQI by 94% at low load, and 79∼93% at medium to high load. Throughput wise, for the realistic arrival rates that are well below the cell capacity, it offers the same average cell throughput and less than 25% degraded cell edge user throughput. Although the reduction is performed only for the CQI in a SIMO system, the overhead of PMI and RI in a MIMO system can also be reduced by using a similar approach.

To reduce the overhead of ACK/NACK signaling, a method called ACK/NACK bundling is applied. The loss of ACK/NACK bundling decreases with the number of users (load) per cell. In a SIMO system with 10 users per cell, having one ACK/NACK per user causes a loss of 7.5% in average cell throughput and 12% in cell edge user throughput. It is also observed that in the considered macro-cell scenario with 2 by 2 MIMO and full buffer traffic model, the number of users supporting rank-2 transmission is rather low, and hence bundling the ACK/NACKs of the MIMO streams (intra-CC bundling) gives no performance loss. However, for cases with a high probability of rank-2 transmissions, or higher MIMO ranks, the throughput loss due to intra-CC bundling may increase. Further investigations are needed in order to decide the number of ACK/NACKs in the antenna domain.

The impact of overhead reduction on the uplink coverage is also analyzed based on some simplified assumptions. The final decision of the feedback method for a user should be made according to the cell load as well as the user uplink capability.

Chapter 6

Conclusions and Future Work

The thesis has investigated system level technologies to improve the performance of the Long Term Evolution (LTE)-Advanced systems. LTE-Advanced should support a wide bandwidth of up to 100 MHz, which is usually obtained by aggregating several Component Carrier (CC)s.

Both local area and wide area scenarios have been considered in this thesis. The reason for considering the wide area is that this is the way the traditional wireless networks are deployed, and it will also be one of the most important scenarios for the future systems. As to local area network, it has a different interference condition from the wide area network, due to the small cell size and the uncoordinated placement of the Access Point (AP)s. Furthermore, it usually lacks a central controller and hence prefers the self-organized operation. In view of these aspects, the two scenarios have been studied separately in this thesis, and different approaches have been used for them. Recommendations for the LTE-Advanced system in local and wide area networks are provided in the following sections. The possible future work is listed at the end of the chapter.

6.1 Recommendations for Local Area Network

Chapter 2 investigates the performance of a local area TDD system with different levels of network time synchronization. It is observed that the time synchronization

has a significant impact on the system performance. Without synchronization, the uplink transmission will be severely penalized by the interference coming from the downlink. In order to guarantee reasonable uplink service, at least loose network time synchronization should be achieved.

The other study carried out in local area is the performance analysis of frequency reuse, and the development of distributed algorithms for spectrum sharing in unplanned scenarios. The results in Section 3.2 indicate that a reuse factor of 2 is preferable in most cases. Afterwards, several algorithms have been developed to facilitate the spectrum sharing without the need of a central controller. One of the algorithms is FiDCA, it aims to achieve the optimal reuse-2 pattern by allowing each cell to jump towards the channel with low interference. Another algorithm, the DS^3, reduces interference by preventing cells from operating on spectrum with heavy interference. Compared to the case of reuse-1, they achieve a gain of 50$\sim$500% in cell edge user throughput, with small or no loss in average cell throughput. The selection between these algorithms in a specific local area network can be done depending on the variation range of the cell density and the processing capability of the AP.

Achieving network time synchronization is also needed in the wide area network, which has already been extensively studied. Furthermore, the finding from the spectrum sharing study also confirms the advantage of having frequency reuse-1 in the wide area.

6.2 Recommendations for Wide Area Network

The first topic investigated for the wide area network is the carrier load balancing. It is found that the multi-user diversity is maximized when the load is strictly balanced across the CCs. Among several existing techniques, the balancing method of Least Load (LL), which assigns the CC with the lowest load to a user, offers the best performance. To provide fairness among users that can access different number of CCs, a cross-CC PF scheduler is proposed. It increases the scheduling priority for the users with less CCs and hence achieves much higher cell edge user throughput than independent scheduling. The gain is up to 90% with full buffer transmission and 40% with finite buffer transmission. Meanwhile, it has no degradation in the average cell throughput, and is therefore an attractive candidate for the layer-2 packet scheduling.

The last topic of this study is the feedback reduction, which is carried out for two kinds of uplink information: the CQI, which is used for packet scheduling and link adaptation; and the ACK/NACK of the packet receptions. The CQI can be reduced both within each CC and across the CCs. For the intra-CC reduction,

two known methods of average best-M and wideband CQI have been investigated with full CQI feedback as the reference. It is found that wideband CQI is efficient when there are few users per cell. Otherwise, best-M is recommended to offer the frequency domain diversity. For the inter-CC reduction, a load adaptive CQI reduction scheme is proposed. It assigns less CCs to the LTE-Advanced users when the load is high, and vice versa. The proposed scheme also includes the selection between wideband CQI and average best-M. It significantly reduces the overhead (79∼94%) while retaining reasonable downlink performance. The ACK/NACK overhead is reduced by bundling the multiple ACK/NACKs of a user. It also offers the trade-off between downlink performance and uplink overhead.

The investigated topics in the wide area are in general applicable also for the local area network, and similar behavior is expected.

6.3 Future Work

In this thesis the local and wide area networks occupy different frequency bands and therefore their performance has been analyzed separately. However, due the spectrum scarcity, they may be deployed on the same frequency band and the interaction between these two kinds of networks is significant. Therefore, it is necessary to study the case when both of them exist on the same band, e.g., the heterogeneous networks.

The investigations in the multi-CC system are mostly for the case when the aggregated CCs have the same channel quality and the same center carrier frequency, and users are not subject to any Quality of Service (QoS) constraint. It is necessary to study the performance of quality guaranteed services in combination with different carrier aggregation patterns, where the CCs can differ from each other in bandwidth, carrier frequency, interference level, etc.

Several techniques have been investigated in this thesis to increase the average cell throughput and cell edge user throughput, or to trade off between these two performance metrics, including Generalized Proportional Fair (GPF) scheduling, hard or fractional frequency reuse, load-adaptive average best-M, velocity dependent CC assignment. However, there are many others that should be considered. For instance, MIMO is a very important technique to boost the system performance. Up to 8×8 MIMO in the downlink and 4×4 in the uplink have been approved in 3GPP for the LTE-Advanced system. How to optimize the MIMO transmission with the existence of multiple CCs should be investigated. When multi-user MIMO is used, the transmissions can only be multiplexed for users that are assigned on the same CC. In this case, further investigation is needed to divide the users among the available CCs. In the uplink transmission, operating

across a wide bandwidth requires high transmit power, which may be problematic for users far from the base station. Therefore, the power control should also be studied with the spectrum/CC assignment.

Appendix A

Scenario Dependent Simulation Assumptions

This appendix describes the simulations assumptions for each considered scenario in this thesis, including the indoor home and office scenarios (local area) and the outdoor Manhattan, based on the proposal in [113], and the wide area Macrocell scenario according to [114]. The system parameters are taken from the LTE system, which are likely to be used for LTE-Advanced as well.

A.1 Local Area Network

A.1.1 Home Scenario

A local area home scenario with 16 homes is considered. Each home contains 4 rooms, and one room has a size of 5×5 m^2. There is one AP in each home. An AP is a low transmit power base station, as compared to the ones used in wide area networks. The house owner is completely free to place the AP wherever possible, leading to randomized AP positions. It is reasonable to assume that a user is served by its own AP, thereby receiving service from its own AP even if the signal strength from a neighboring AP is higher. This kind of cell selection is known as the CSG [115, 116].

Fig. A.1 gives an example of a home scenario with randomized APs (denoted by the solid circles). Users are randomly distributed within the coverage of each cell. Neighboring cells/homes are distinguished from each other by different colors. While the walls inside each home are thin walls with small attenuation loss, the ones between homes are thick, and the attenuation loss is high. The high attenuation loss to some extent offers 'isolation' of the individual home and helps to improve the SINR condition.

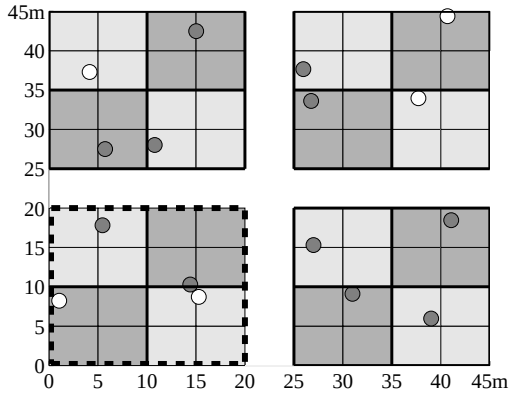


Fig. A.1: Indoor home scenario with 16 cells.

Due to the small number of users that are served by each local area AP, the traffic changes very fast within each cell and quite often, one cell will be in idle mode as users connected to it do not demand any data traffic. In the snapshot based simulation, the probability of having traffic in a cell is controlled by a pre-defined activity factor. In the long term, this activity factor becomes equivalent to the ratio of the active APs over the total number of APs, and it also indicates the cell density. When the total number of base stations is fixed, a high activity factor means that more cells are active, and the density is high. In Fig. A.1 and Fig. A.2, a solid circle means an active AP, whereas an empty circle indicates an idle AP. To verify the applicability of the findings with different network sizes, a small home scenario is also considered for the performance evaluation. It contains only 4 homes in the bottom left corner of Fig. A.1, and all APs are assumed to be active.

A.1.2 Office scenarios

In the office scenario, the location of APs can be pre-planned so that each will cover a certain area. Fig. A.2 shows an office scenario with 16 cells. The APs

are placed at the center of each cell, and users are randomly distributed within the whole area. Unlike the home scenario, which restricts a user to be served by the AP physically located within the same cell, the users in an office scenario are able to connect to the AP with the highest channel gain (path loss and shadowing fading). This is often referred to as the OSG mode [117]. The office scenario considered in this study is constituted of 16 equal-size cells, where each cell has 10 rooms, separated into two rows by a corridor of 5 m width. The size of each room is 10×10 m^2. The office scenario is constructed with only thin walls so the cell isolation is worse than the home scenario. However, the coordinated AP placement and the OSG access mode prevents a user from being connected to an AP with poor channel quality while interference is high, hence better cell edge SINR is achieved than in the home scenario. Similar as with the home scenario, a small office scenario with only 4 cells is also considered for the performance evaluation.

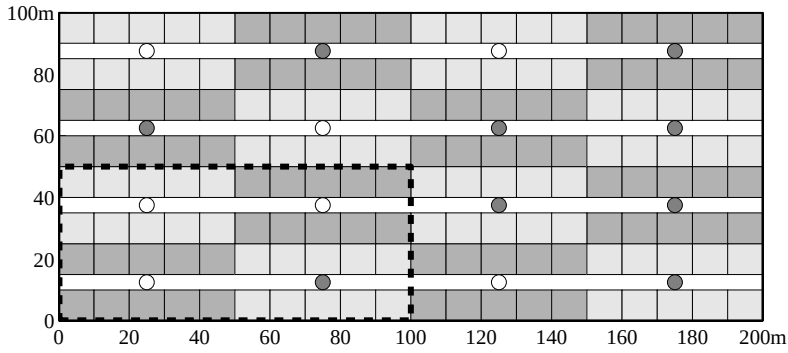


Fig. A.2: Indoor office scenario with 16 cells.

A.1.3 Manhattan Scenario

The Manhattan scenario is used to verify the advantage of frequency reuse-1 in wide area networks. Following [18, 113], the locations of the base stations are coordinated such that Line of Sight (LOS) connection between neighboring base stations is avoided. Similar as in the office scenario, users are connected to the base station with the highest path gain, i.e., the Open Subscriber Group (OSG) mode. This scenario is shown in Fig. A.3. Each block means one building of size 110×110 m^2; the road has a width of 15 m.

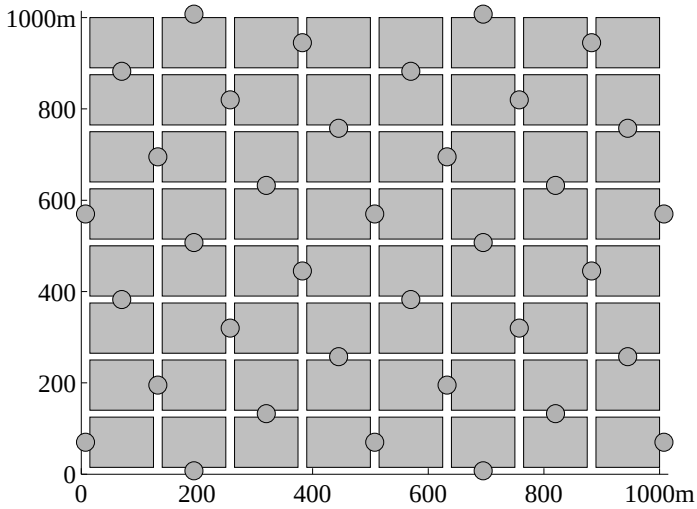


Fig. A.3: Manhattan scenario with 36 cells and coordinated base station locations.

A.1.4 General Assumptions of LTE-Advanced in Local Area and the Manhattan Scenario

Table A.1 gives a short summary of the system assumptions for the considered scenarios.

The snapshot based simulation is used for the performance evaluation. Within each snapshot, the base stations are generated according to the deployment method (at the center of each cell or randomly located) as well as the activity factor. A random number of users is then generated within each active cell, which uniformly distributes within the range of 2 to 4 in the home scenario and 5 to 10 in the office scenario. In order to get the statistically averaged performance, a few thousands of snapshots are simulated and the throughput is collected from all these snapshots. The Manhattan has a much larger cell coverage than the indoor networks, therefore it has different system assumptions. To be specific, 10 outdoor users are simulated within each cell, and the downlink transmit power is 30 dBm. The location & distance dependent path loss model (B1) is used. Because the outer-tier cells do not have realistic interfering cells in the simulation, the results are collected only from the center 7 cells out of the 36 simulated ones. This is different from the indoor simulations, where all simulated cells contribute to the overall performance. In the Manhattan scenario, the path loss is calculated according to Table A.2 [18, 113].

In Table A.2, f_c is the center carrier frequency in GHz; $h'_{BS} = h_{BS} - 1.0$, $h'_{MS} =$

Table A.1: Parameters and assumptions for system level evaluation of local area networks [18, 113].

Parameter	Setting/description
Spectrum allocation	100 MHz at 3.5 GHz
Duplexing scheme	TDD
Users per cell	Home: 2∼4 users; Office: 5∼10 users; Manhattan: 10 users
Transmit power	AP: 24 dBm in home and office scenarios, 30 dBm in Manhattan scenario; user: -30∼24 dBm
Antenna system	"Omni-directional". AP: 3 dBi; user: 0 dBi
Receiver noise figure	AP: 7∼9 dB; user: 9 dB
Room size	Home: 5×5 m^2; Office: 10×10 m^2
Corridor width	5 m
Indoor wall attenuation loss	Inside home: 5 dB; between homes: 10 dB Office: 5 dB
Indoor path loss	LOS: $18.7log_{10}d + 46.8 + 20log_{10}(fc/5.0)$ NLOS: $20log_{10}d + 46.4 + n_w L_w + 20log_{10}(fc/5.0)$ where d = direct-line distance [m] f_c = carrier frequency [GHz] n_w = number of walls between transmitter and receiver L_w = wall attenuation [dB]
Shadow fading	Uncorrelated, log-normal with 3 dB std.
Minimum Coupling Loss	45 dB
Traffic model	Full buffer

Table A.2: Location & distance dependent path loss model for the outdoor Manhattan scenario [18, 113].

Scenario	Path loss [dB]	Shadow fading	Applicable range
LOS	$PL = 22.7log_{10}(d_1) + 41.0 + 20log_{10}(f_c/5)$	3 dB std.	$d_1 < d_{BP}$
	$PL = 40.0log_{10}(d_1) + 9.45 - 17.3log_{10}(h'_{BS}) - 17.3log_{10}(h'_{ms}) + 2.7log_{10}(f_c/5.0)$		$d_1 \geq d_{BP}$
NLOS	$PL = min\{PL(d_1, d_2), PL(d_2, d_1)\}$ where $PL(d_k, d_l) = PL_{LOS}(d_k) + 20 - 12.5n_j + 10n_j log_{10}(d_l) + 3log_{10}(f_c/5.0)$ and $n_j = max\{2.8 - 0.0024d_k, 1.84\}$ $k, l \in \{1, 2\}$ PL_{LOS} is the pathloss of the LOS scenario.	4 dB std.	$d_1 < 5$km

$h_{MS} - 1.0$, where $h_{BS} = 10$ m, $h_{MS} = 1.5$ m are the actual antenna heights of the base station and the user equipment; the effective environment height in urban environments is assumed to be 1.0 m. $d_{BP} = 4h'_{BS}h'_{MS}f_c[\text{Hz}]/c$, with $c = 3.0 \times 10^8$ m/s being the propagation velocity in free space.

The Non-Line of Sight (NLOS) path loss model for the outdoor Manhattan scenario is dependent on two distances, d_1 and d_2. These distances are defined with respect to a rectangular street grid, as illustrated in Fig. A.4. d_1 is the distance from the base station to the center of the perpendicular street, and d_2 is the distance of the user along the perpendicular street, measured from the center of the horizontal street where the base station is installed [18, 113].

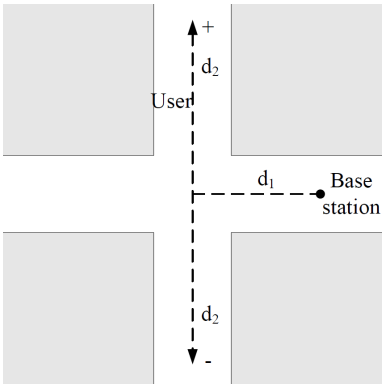


Fig. A.4: Geometry for the location & distance dependent path loss model.

A.1.5 The Physical Layer Modeling

The performance with different synchronization levels is investigated by system level simulations, using a snapshot based method. One radio frame with 10 ms duration is simulated for each snapshot, and the effect of the uplink-downlink mutual interference due to imperfect network time synchronization can be seen from the different interference levels for the individual subframes. Within each snapshot, the cell layout is generated according to the scenario. For the office scenario, one AP is placed at the center of each cell, but for the home scenario, APs are randomly placed. Users are uniformly distributed inside each cell. The Error Vector Magnitude (EVM) is used to model the hardware imperfection [22], which is assumed to have a value of 3%. The use of EVM puts a maximum limitation on the received SINR, which is:

$$SINR_{max} = -20log_{10}0.03 = 30.5dB \tag{A.1}$$

The received SINR then becomes [118]:

$$SINR = 1/\left(1/SINR_{ideal} + 1/SINR_{max}\right) \tag{A.2}$$

In (A.2), the ideal SINR is calculated from the signal power, the interference power and the noise level. The LTE link-level capacity curve is used to map the SINR to spectral efficiency and finally the throughput. According to [68], the capacity in a Single Input Single Output (SISO) LTE system can be estimated by:

$$S = min\left\{BW_{eff} log_2(1 + SINR/SINR_{eff}), 5.02\right\} \tag{A.3}$$

where S is the estimated spectral efficiency in bps/Hz, which is upper limited according to the hard spectral efficiency given by 64-Quadrature Amplitude Modulation (QAM) with the coding rate 9/10 (5.4 bps/Hz) and the cyclic prefix (0.93); BW_{eff} adjusts for the system bandwidth efficiency and $SINR_{eff}$ adjusts for the SINR implementation efficiency. $\{BW_{eff}, SINR_{eff}\}$ has the value of $\{0.56, 2.0\}$ in the downlink, and $\{0.52, 2.34\}$ in the uplink. Although (A.3) treats the interference as Gaussian, it has been shown in [68] to provide an accurate throughput estimation in a multi-cell system with non-Gaussian interference.

According to the 3GPP standard [22], the local area AP transmits at a constant power of 24 dBm in the downlink. Note that for the sake of simplicity, the same power level is used for both home and office scenario, despite the fact that a lower transmit power level of 20 dBm is specified for the home scenario. In the uplink transmission, Power control is applied to minimize the interference as well as to prolong the user battery life [119]. The uplink power control mechanism contains both open-loop and closed-loop operations. Specifically, the open-loop component partially compensates for the pass loss [120], whereas the closed-loop component offers adaptation to the channel quality variations [121–123]. Because the channel varies slowly in the indoor system, fast fading is not modeled and fractional path loss compensated power control is applied. The transmit power for a user is [27, 119]:

$$P = max\left\{P_{min}, min\{P_{max}, P_0 + \alpha PL\}\right\} \tag{A.4}$$

where $P_{min} = -30$ dBm, $P_{max} = 24$ dBm, PL is the path loss value in decibel between a user and its serving AP. In [124], it was found that $P_0 = -60$ dBm and $\alpha = 0.8$ give a good performance in local area networks. These values are adopted here for the uplink power control.

To simplify the problem and focus mainly on the effect of network time synchronization and cell level spectrum sharing, a frequency domain channel blind Fair Resource (FR) scheduling is used within each cell. By using FR, each user gets an equal share of the available resources of its own cell. Full buffer traffic model is used, which assumes each user always has data to transmit or receive. As a consequence, all possible transmission resources would be used. Furthermore, only LTE-Advanced users are considered, which can access the whole spectrum.

The following metrics are used for the evaluation of the system performance:

Average cell throughput: the cell throughput averaged over all simulated cells.

Cell edge user throughput: the 5% worst user throughput from all simulated ones.

A.2 Wide Area Network

The performance of the load balancing methods, the different PS algorithms, and the feedback reduction is evaluated in a quasi static downlink system level simulator that follows the LTE-Advanced specifications defined in [9], including layer-3 CC selection, layer-2 PS, HARQ, link adaptation and feedback functionalities. The simulation scenario is 3GPP Macro-cell case#1 as defined in [114, 125]. The simulation parameters are summarized in Table A.3. The detailed assumptions are discussed in the following.

A.2.1 Macro-cell Layout and Sectorization Within Each Site

The macro-cell case#1 uses hexagonal site layout and an inter site distance of 500 m. A three-sector antenna is mounted in each site, which sectorizes one site into three cells. The 2-dimension antenna pattern is simulated, as [114]:

$$A(\theta) = -min\left[12\left(\frac{\theta}{\theta_{3dB}}\right)^2, A_m\right] \tag{A.5}$$

where $-180 \leq \theta \leq 180$ degree, $\theta_{3dB} = 70$ degree is the 3 dB beamwidth when the antenna gain is reduced to half the maximum value, $A_m = 20$ dB is the maximum attenuation.

Among the 19 macro-sites, only the center-site is simulated. The surrounding sites are used to generate time continuous interference across the full bandwidth. The accuracy of the center-site simulation as compared to simulating all sites (with wrap-around) is verified by the G-factor distribution presented in Fig. A.5. G-factor is the ratio of the total received wideband signal power and the interference plus noise power at the receiver side. It includes the effects of path loss and shadow fading, but is average over fast fading. It is equivalent to the average wideband SINR in a single antenna system.

The layout of the 19 hexagonal sites is depicted in Fig. A.6(a), and the three-sector antenna pattern within each site is shown in Fig. A.6(b).

Table A.3: Parameters and assumptions for system level evaluation of wide area LTE-Advanced networks.

Parameter	Setting/description
Test scenario	Macro-cell case#1 (19 sites, 3 cells per site)
Carrier aggregation pattern	4 CCs at 2.0 GHz frequency; 10 MHz per CC
Number of PRBs per CC	50 (12 sub-carriers per PRB)
Subframe duration	1 ms (11 data symbols plus 3 control symbols)
Modulation and coding	QPSK (1/5, 1/4, 1/3, 2/5, 1/2, 3/5, 2/3, 3/4), 16-QAM (2/5, 9/20, 1/2, 11/20, 3/5, 2/3, 3/4, 4/5, 5/6), 64-QAM (3/5, 5/8, 2/3, 17/24, 3/4, 4/5, 5/6, 7/8, 9/10)
Base station transmit power	46 dBm per 10 MHz
Antenna system	base station: 14 dBi gain, one antenna; user: 0 dBi gain, 2 antennas with Interference Rejection Combining
1^{st} transmission BLER target	10%
HARQ modeling	Ideal Chase Combining with maximum 4 transmissions
CQI reporting	1 CQI per 3 PRBs; 1.6 dB quantization step; log-normal error with 1 dB std; 6 ms delay
Layer-2 packet scheduling	RR in time domain, PF in frequency domain
Admission control constraint	Maximum 50 users per cell
User distribution	Uniform
Path loss	$128.1 + 37.6 \log_{10}(d[km])$
Shadow fading	Log-normal with 8 dB std; full correlation within each site; correlation coefficient of 0.5 between sites
Minimum distance between user and base station	35 m
User speed	3 kmph
Receiver noise figure	9 dB (-105 dBm/MHz)
Access mode	Path loss based, OSG
Traffic type	Full buffer (with 10 users) and finite buffer (Poisson arrival with buffer size of 2 Mbits)

A.2.2 Wireless Channel Modeling

For the quasi static simulation, the location of each user is fixed during the simulation period, i.e., the path loss and shadow fading remain constant. However, the user is assumed to have a velocity of 3 kmph and hence experiences fast fading. The path loss is calculated as [114]:

$$PL = 128.1 + 37.6 \log_{10}(d[km]) \tag{A.6}$$

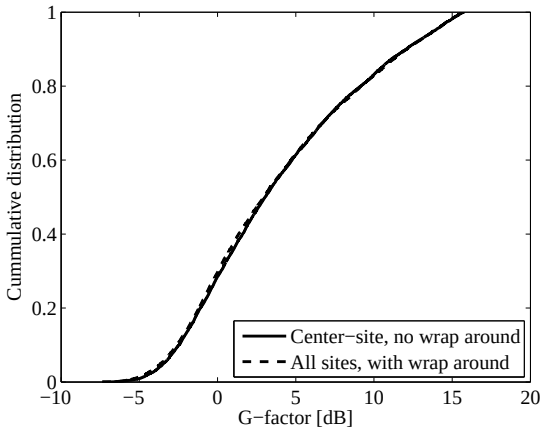
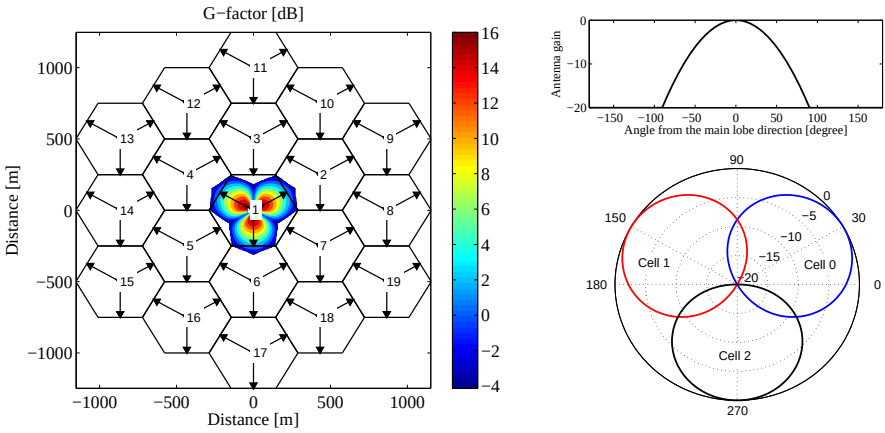


Fig. A.5: G-factor distribution for the center-site and all-site simulation modes.



(a) 19-site macro-cell layout

(b) 3-sector antenna pattern

Fig. A.6: The 19-site macro-cell layout and 3-sector antenna pattern.

All the users are modeled as indoor users, therefore an additional outdoor-to-indoor penetration loss of 20 dB is added on top of the path loss calculated using (A.6).

The shadow fading is assumed to follow a log-normal distribution with 8 dB standard deviation. While fully correlated shadow fading is considered within each site, the shadow fading between sites is spatially correlated, with a correlation

coefficient of 0.5. The correlated shadow fading between user k and site i is:

$$S_{k,i} = \sqrt{\rho} S_k + \sqrt{1-\rho}\sigma n_i \tag{A.7}$$

In (A.7), $\rho = 0.5$ is the site correlation; $\sigma = 8$ dB is the shadow fading standard deviation; $S_k = \sigma n_k$ is the near-field shadow fading for user k; n_k and n_i are random variables that follow the normal distribution.

The multi-path channel is modeled using the tapped-delay model. Assuming a finite number of P distinct paths, the channel impulse response can be written as [12]:

$$h(\tau) = \sum_{i=1}^{P} a_i \delta(\tau_i) \tag{A.8}$$

In (A.8), each tap is the superposition of a large number of scattered waves with approximately the same delay of τ_i. a_i is time varying due to the multi-path phenomenon and the power for each tap is $\left|a_i^2\right|$. The amplitude of a_i is usually assumed to be Rayleigh distributed and with Doppler spectrum [126]:

$$S(f) \propto 1/(1-(f/f_D)^2)^{0.5} \tag{A.9}$$

where $f_D = v/\lambda = vf_c/c$ is the maximum Doppler shift, v is the mobile speed in m/s, $c = 3.0 \times 10^8$ m/s is the speed of light, f_c is the carrier frequency in Hz. In the 2 GHz band with user speed of 3 kmph, the maximum Doppler shift is 5.6 Hz.

In this thesis, the Typical Urban model with 20 paths (TU-20) [126] is used. The power delay profile for the TU-20 model is given in Table A.4.

A.2.3 Link Adaptation

The link to system mapping is based on the Exponential Effective SIR Mapping (EESM) model [127]. It calculates an effective SINR based on the individual SINR value within each PRB.

$$SINR = -\beta \ln\left(\frac{1}{N_u}\sum_{k=1}^{N_u} e^{-\frac{SINR_k}{\beta}}\right) \tag{A.10}$$

In (A.10), β is a parameter optimized from link-level simulation results for each Modulation and Coding Scheme (MCS). The value of β also indicates the sensitivity of turbo decoder with respect to channel variation. In general, a high level of MCS is more sensitive to the channel variation than a low MCS and hence requires a higher value of β. N_u is the number of sub-carriers within one PRB. The SINR value in (A.10) is also subject to EVM, which is assumed to be 8%, corresponding to a maximum SINR of 22 dB.

Table A.4: Power delay profile for TU-20 channel.

Tap number	Tap delay [μs]	Relative tap power [dB]
1	0	-5.7
2	0.217	-7.6
3	0.512	-10.1
4	0.514	-10.2
5	0.517	-10.2
6	0.674	-11.5
7	0.882	-13.4
8	1.230	-16.3
9	1.287	-16.9
10	1.311	-17.1
11	1.349	-17.4
12	1.533	-19.0
13	1.535	-19.0
14	1.622	-19.8
15	1.818	-21.5
16	1.836	-21.6
17	1.884	-22.1
18	1.943	-22.6
19	2.048	-23.5
20	2.140	-24.3

Outer-Loop Link Adaptation (OLLA) is used to stabilize the 1^{st} transmission BLER as a practical remedy against link adaptation errors due to inaccurate SINR reports [109, 128]. If an ACK is received for the first transmission, the OLLA offset will be decreased by Δ_{down} dB. Otherwise, it will be raised by Δ_{up} dB. The relation between Δ_{down} and Δ_{down} is:

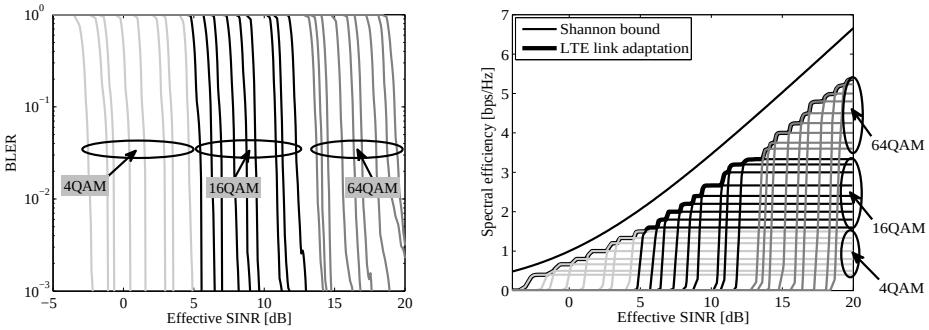
$$\Delta_{down} = \Delta_{up} \frac{BLER}{1 - BLER} \tag{A.11}$$

In this study, the step of Δ_{up} is 0.5 dB, and with 10% BLER target, $\Delta_{down} = 0.0556$ dB. The OLLA offset is also restricted within a range of -3$\sim$5 dB. During the link adaptation process, the OLLA offset is added on top of the feedback SINR value for selecting the proper MCS.

Following the LTE specification, three modulation levels of Quadrature Phase Shift Keying (QPSK), 16-QAM and 64-QAM have been used. Together with turbo-coding, in total 26 MCSs have been considered. The SINR to BLER mapping for these MCSs is summarized in Fig. A.7(a), and their achievable spectral efficiency at different SINR values is shown in Fig. A.7(b). The Shannon bound and the performance with link adaptation are also plotted, assuming a BLER target of 10%. The SINR threshold and β value for EESM of each MCS is summarized in

Table A.5, which is used in the system level simulation.

The specific curves are generated by using Turbo receiver in TU-20 channel, when each user is statically scheduled with 26 PRBs (312 sub-carriers), and coding is performed within these PRBs. According to link level simulation, it is enough to achieve maximum interleaving gain by coding across these PRBs. For cases when a user is scheduled with less number of PRBs, this physical layer mapping is a bit optimistic. The accuracy in general is acceptable because the packet scheduler usually does not schedule more than 6 or 7 users in a 10 MHz bandwidth within each TTI.



(a) SINR versus BLER

(b) SINR versus spectral efficiency

Fig. A.7: Mapping from SINR to BLER and spectral efficiency.

A.2.4 Hybrid Automatic Repeat Request Management

LTE(-Advanced) makes use of HARQ to ensure success packet reception at the receiver side. Re-transmission will take place when the previous transmission fails. Asynchronous and adaptive HARQ is supported in LTE(-Advanced), which means the re-transmission can be scheduled at any time on any PRBs. However, the same MCS should be used, such that the re-transmission can be combined with the previous transmission(s). In this study, the Chase Combining is used, and the SINR after Chase Combining is modeled as:

$$SINR_n = \epsilon^{n-1} \sum_{k=1}^{n} SINR_k \tag{A.12}$$

In (A.12), $SINR_k$ is the received SINR for the k^{th} (re-)transmission; $SINR_n$ is the combined SINR after n transmissions, and ϵ quantizes the combining effi-

Table A.5: Link to system level mapping table.

Modulation	Coding	Spectral Efficiency [bps/Hz]	10% BLER Threshold [dB]	EESM-β
QPSK	1/5	0.4	-2.7	1.34
	1/4	0.5	-1.3	1.57
	1/3	0.67	-0.8	1.44
	2/5	0.8	-0.2	1.52
	1/2	1.0	1.3	1.53
	3/5	1.2	2.7	1.58
	2/3	1.33	3.4	1.58
	3/4	1.5	4.6	1.69
16-QAM	2/5	1.6	5.3	4.33
	9/20	1.8	6.2	4.75
	1/2	2.0	6.8	5.25
	11/20	2.2	7.8	6.08
	3/5	2.4	8.7	6.64
	2/3	2.67	9.3	6.69
	3/4	3.0	10.7	7.45
	4/5	3.2	11.2	7.40
	5/6	3.33	12.2	8.10
64-QAM	3/5	3.6	13.6	21.25
	5/8	3.75	14.0	22.62
	2/3	4.0	14.5	22.91
	17/24	4.25	15.4	25.80
	3/4	4.5	16.3	28.74
	4/5	4.8	16.8	29.29
	5/6	5.0	17.8	31.07
	7/8	5.25	18.6	33.15
	9/10	5.4	19.2	33.78

ciency. Ideal Chase Combining is used, therefore, $\epsilon = 1$. The maximum number of transmissions is $n = 4$ in the study.

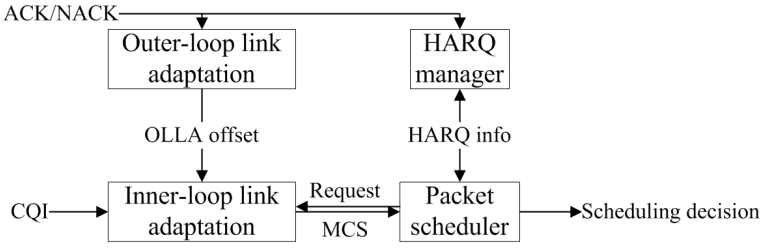


Fig. A.8: Interaction between link adaptation, HARQ and packet scheduling.

Fig. A.8 summarizes the interaction between OLLA, Inner-Loop Link Adaptation (ILLA), HARQ management and packet scheduling. As shown, the ACK/NACK information is used to adjust the OLLA offset, which is added on top of the CQI feedback and used in ILLA. ACK/NACK also affects the HARQ management process and in turn, the packet scheduling (by restricting the re-transmission to use the same MCS as the previous one). After assigning resources to the users, the packet scheduler will request ILLA to make the final decision on the MCS to each user. The HARQ manager needs to update the stack information based on the scheduling decision.

A.2.5 Interference Rejection Combining

In a wireless communication system with fading channel, using multiple antennas at the receiver side can exploit the spatial diversity to improve the signal quality. When the noise at the different receiving antenna branches is uncorrelated, the best combining method is Maximal Ratio Combining (MRC), which requires the knowledge of the channel gain and the noise power per branch. However, the cellular system is usually interference limited, and the interferences at the receiving antenna branches are not fully uncorrelated. In this situation it is beneficial to use an Interference Rejection Combining (IRC) receiver to reject the dominating interference [129].

With one transmitting antenna and L receiving antennas, the received signal is:

$$r = hx + n \tag{A.13}$$

For a system with L receiving antennas and I interferers, $h_{l,i}$ is the complex channel gain between the l^{th} receiving antenna and the i^{th} interferer. The received interference-plus-noise at the l^{th} antenna branch is:

$$n_l = v_l + \sum_{i=1}^{I} P_i h_{l,i} \tag{A.14}$$

where v_l is the thermal noise component with variance σ_v^2, which is equal for all antennas; P_i is the power of the i^{th} interference.

The combining process in the receiver can be represented by:

$$y = w^T r = w^T hx + w^T n \tag{A.15}$$

In (A.15), w^T is a weighting vector. Different w^T corresponds to different combining technique. After combining, the SINR can be expressed as [130]:

$$SINR = \frac{E[w^T hh^H w^* |x|^2]}{E[w^T nn^H w*]} = \frac{w^T hh^H w^* P_s}{w^T R w^*} \tag{A.16}$$

where P_s is the received signal power and $R = E[nn^H]$ is the noise-plus-interference covariance matrix. In the case with two receiving antennas, R has the form of [130]:

$$R = \begin{bmatrix} \rho_1^2 & \rho_{1,2} \\ \rho_{1,2}^* & \rho_2^2 \end{bmatrix} \tag{A.17}$$

where $\rho_l^2 = \sigma_v^2 + \sum_{i=1}^{I} P_i |h_{l,i}|^2$ is the received interference-plus-noise power, and $\rho_{1,2} = \sum_{i=1}^{I} P_i h_{1,i} h_{2,i}^*$.

For the IRC receiver, a weighting vector of $w^T = h^H R^{-1}$ is used. By inserting this into (A.16), the output SINR of an IRC receiver is [130]:

$$SINR_{IRC} = h^H R^{-1} h P_s \tag{A.18}$$

A.2.6 Traffic Model

Both full buffer and finite buffer traffic models are considered in the simulation, according to the guideline in [131]. The full buffer traffic model is characterized by the fact that each user always has data to transmit or receive, and the transmission never finishes during a simulation run. The simulation process with the full buffer traffic is conducted as a series of simulation runs (5 seconds in each run) with a constant number of users per cell. Different cell load can be achieved by varying the number of users per cell. Many simulation runs are needed to cover the user locations within each cell. Although not very realistic in many cases, the full buffer traffic is extensively used because of the simplicity.

The finite buffer traffic model is characterized by the fact that the users have a finite amount of data to transmit or receive. Once the data transmission is finished, the user is removed from the system. For the finite buffer model with dynamic birth-death, only one long simulation run (with duration of hundreds of seconds) is launched. This is valid because users are created and terminated dynamically during the simulation. In this study the arrival of the users is modeled as Poisson arrival, and the data associated with different users are assumed to be of the same size. By changing the Poisson arrival rate, different load conditions can be obtained.

Appendix B

Capacity of MIMO Systems and the Multi-user Gain

B.1 Introduction

The capacity of a wireless system can be improved in several ways, e.g., adjust the transmission rates according to the channel quality using link adaptation [25]; use MIMO multiplexing to send multiple data streams or obtain MIMO diversity gain [132,133]. In a multi-user system, the system capacity can be further improved by exploiting the multi-user diversity using channel aware packet scheduling [24, 104]. These techniques have been extensively used in many existing systems, e.g., WiMAX [134], the 3GPP LTE [10]. They will also be applied in the forthcoming IMT-Advanced systems [46]. It is therefore worthwhile to analyze the performance combining all these techniques.

There have been several attempts to present analytical expressions for the system performance in this case. Starting with references [135–139], they considered one or a combination of two of the above mentioned techniques. Reference [140] addressed all of them, but for spatial multiplexing MIMO systems only. In a real system the capacity is limited by the highest supported modulation and coding scheme, a fact that is usually not considered in analytical derivations such as the ones mentioned above, but have been included in simulation based analyses, e.g., in [97, 141–144]. Simplified approximations for a fixed number of users per cell were presented in [68] for an LTE system under various MIMO configurations and scheduling methods,

based on simulations.

In this appendix, the exact closed form expression of the MIMO capacity in terms of average spectral efficiency is derived. Similar to [135–140], the derivation is carried out under the assumption of uncorrelated Rayleigh fading channels and the same average channel quality for the multiple users. Both channel blind RR and channel aware PF [104] schedulers are investigated in the capacity analysis. The following antenna configurations are considered: the first one has one transmit and multiple receive antennas, with MRC at the receiver. This is the same assumption as in [136, 137], however the analysis in [137] was carried out with only 3 receiving antennas, and did not address the problem in combination with channel aware packet scheduling. The second case assumes the same number, M, of transmit and receive antennas and Zero Forcing (ZF) precoding as in e.g. [137, 140]. ZF precoding with multi-stream transmission and link adaptation for each data stream offers an M-fold improvement of the peak data rate as compared to single stream transmission. When multiple users exist in a cell, the M data streams may not necessarily come from a single user, but several users can be scheduled at a time, and each user can be allocated on multiple antennas [140]. When the channel quality is poor, multiplexing multiple data streams cannot bring any performance gain. In this situation, it is preferable to transmit with less number of data streams to obtain diversity gain and array gain, i.e., MIMO diversity. This is known as rank adaptation, where 'rank' is the number of transmitted data streams. In this appendix, the capacity expressions for the 2×2 MIMO with rank adaptation are also derived.

From the obtained results, it is noticed that the multi-user gain can be well approximated by a logarithm function of the number of users. For the purpose of modeling, a summary of the approximated gain in various scenarios is given at the end of the appendix.

The rest of this appendix is organized as follows: Section B.2 describes the system model, including link adaptation, the physical link, and packet scheduling. Section B.3 provides the theoretical analysis and derives the exact closed form capacity expressions. In Section B.4, the performance for selected cases is shown. The multi-user gain in an LTE system in both Micro- and Macro-cell scenarios is presented in Section B.5. Finally, Section B.6 concludes the appendix.

B.2 System Model

This appendix focuses on the downlink transmission where multiple users are served by the same base station with a constant transmit power. Fig. B.1 gives an example of such a system. As indicated in the figure, each user first has to esti-

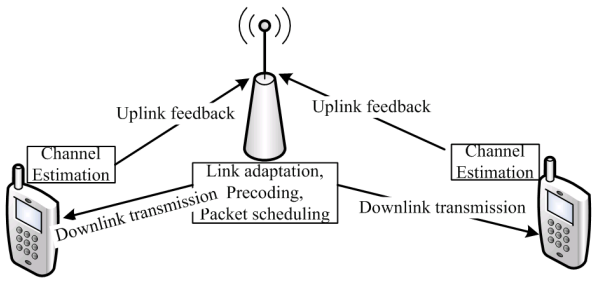


Fig. B.1: Downlink transmission in a multi-user system with uplink feedback.

mate the Channel State Information (CSI) including the SINR and the fast fading channel (only when channel aware precoding is used), then reports to the base station in the uplink feedback channel. The feedback is assumed to be perfect. Based on the feedback information, the base station will format the source bits in the proper way using link adaptation and precoding, and schedule a user with the proper resources. These techniques are described in detail in the following.

B.2.1 Link Adaptation

Link adaptation requires the selection of a proper modulation and coding scheme according to the channel quality which is usually indicated by the SINR value. In practical systems, due to the discrete rate levels and system imperfections, the achievable transmission rate is below the Shannon capacity and upper bounded by the highest modulation and coding scheme. It has been shown that the effect of system imperfections can be well approximated by a constant shift in the SINR values in the decibel scale, or by a weighting factor in the linear domain. Similar to [68, 137, 140], an effective SINR scaling factor is introduced to approximate the performance of practical link adaptation. Additionally, the maximum achievable spectral efficiency C_{max} is also considered. With these constraints, the mapping from SINR to spectral efficiency is:

$$S = min\{\log_2(1 + \alpha\gamma), C_{max}\} \tag{B.1}$$

where S is the achievable spectral efficiency in bps/Hz, γ is the SINR value, and α is the effective SINR scaling factor which lies between 0 and 1. According to Fig. A.7(b), the LTE link-level capacity is about 3 dB below the Shannon limit, i.e., $\alpha = 0.5$. In the following analysis, α is made as an input parameter in the capacity expressions so that the results can be generalized to systems with different link-level capacity.

B.2.2 The Physical Layer Transmission

In case a large number of scatters exist, and there is no LOS link between the transmitter and the receiver, the narrowband wireless channel is usually modeled as Rayleigh fading. As a consequence, the received SINR on each radio link is exponentially distributed with mean value of $\bar{\gamma}$. $\bar{\gamma}$ is decided by the transmit signal strength, the large scale channel gain resulting from path loss and shadowing, the inter-cell interference and the thermal noise level.

In a SIMO system with a single transmit antenna and M receive antennas, an MRC receiver combines the received signals, and the output SINR is ideally the summation of the received SINRs. It follows a central chi-square distribution with $2M$ degrees of freedom, with Probability Distribution Function (PDF) [136, 145]:

$$p_\gamma(x) = \frac{1}{\bar{\gamma}^M (M-1)!} x^{M-1} e^{-x/\bar{\gamma}} \tag{B.2}$$

and the CDF is:

$$F_\gamma(x) = 1 - e^{-x/\bar{\gamma}} \sum_{m=0}^{M-1} (x/\bar{\gamma})^m / m! \tag{B.3}$$

If a system has M transmit and receive antennas, ZF precoding can be used to send up to M parallel data streams to a user. According to [137], the SINR per data stream still follows an exponential distribution. However, because the transmit power is equally shared among M transmit antennas, the mean value for the SINR per data stream is down scaled to $\bar{\gamma}/M$. The PDF of the exponentially distributed SINR for each data stream is therefore [145]:

$$p_\gamma(x) = \frac{M}{\bar{\gamma}} e^{-xM/\bar{\gamma}} \tag{B.4}$$

Correspondingly, the CDF is:

$$F_\gamma(x) = 1 - e^{-xM/\bar{\gamma}} \tag{B.5}$$

When $M = 1$, the expressions degenerate to the SINR distribution of a SISO system.

In a 2×2 MIMO system with rank adaptation, there are two possible ranks. Rank-2 is achieved by ZF precoding, and the SINR distribution is characterized by (B.4) and (B.5). If a user does not support rank-2 transmission, e.g., because of the poor channel quality, rank-1 transmission can be used to improve the SINR by exploiting the array gain and diversity gain. The analysis is carried out with the Alamouti based Space-Time Block Code (STBC), which is widely used for the open loop rank-1 transmission. According to [146], the received SINR for a 2×2 MIMO system with STBC is: $\gamma = \frac{\bar{\gamma}}{2} \sum_{i,j=1,2} g_{i,j}$, where $g_{i,j}$ is the fast fading

channel gain between the i^{th} transmit antenna and the j^{th} receive antenna. This is equivalent to the received SINR for a 1×4 MRC system, except for a 3 dB loss in the average signal quality, due to the total power constraint. Therefore, the capacity of the rank-1 transmission in a 2×2 MIMO system with average SINR of $\bar{\gamma}$ is the same as in a 1×4 MRC system with SINR of $\bar{\gamma}/2$.

B.2.3 Packet Scheduling

In a multi-user system, the transmission resource is shared among the existing users. Let I_k be the scheduling indicator, in a system with K users:

$$I_k = \begin{cases} 1, & \text{user } k \text{ is scheduled} \\ 0, & \text{user } k \text{ is not scheduled} \end{cases} \tag{B.6}$$

With an RR scheduler all users get an equal share of the resources, irrespective of their instantaneous channel quality ('Pr' stands for 'Probability'):

$$Pr(I_k = 1 | \gamma_k = x) = Pr(I_k = 1) = 1/K \tag{B.7}$$

The PF scheduler takes the instantaneous and past averaged user data rate into consideration. On each data stream, it schedules resources to a user that maximizes the scheduling metric r_j/R_j, where j is the user index. Because the users are assumed to have the same average channel quality, they should have the same average data rate R_j. Therefore, maximizing the scheduling metric is the same as maximizing the instantaneous data rate, or the instantaneous channel gain. The selection of the proper user k can be formulated as:

$$k = \underset{j}{argmax}\{r_j/R_j\} = \underset{j}{argmax}\{r_j\} = \underset{j}{argmax}\{g'_j\} \tag{B.8}$$

where R_j, r_j and g'_j are the past averaged data rate, the instantaneous data rate and instantaneous channel gain for user j on a given data stream. The probability of one user being selected when the SINR equals x is:

$$Pr(I_k = 1 | \gamma_k = x) = Pr(\gamma_j < x, \forall j \neq k) = F_\gamma(x)^{K-1} \tag{B.9}$$

Using equations (B.2) to (B.5), the average share of resources for one user with PF is: $\int_{x=0}^{\infty} p_\gamma(x) F_\gamma(x)^{K-1} dx = 1/K$ for both SIMO and MIMO systems. Therefore, PF is exactly fair in terms of resource allocation among the multiple users, on condition that they have the same average channel quality.

B.3 Derivation of the System Capacity with Packet Scheduling

Using the multi-antenna system model described in Section B.2, it is possible to derive the system capacity in terms of average spectral efficiency. The capacity in a system with S data streams ($S = 1$ for SIMO, and $S = M$ for MIMO with ZF precoding) is dependent on the number of users in a cell, the instantaneous spectral efficiency limit (B.1), the received SINR distribution (B.2) or (B.4), and the scheduling probability (B.7) or (B.9). It can be expressed as:

$$C(M, K, \alpha, \bar{\gamma}, C_{max}) = \\ KS \int_0^\infty min\{\log_2(1+\alpha x), C_{max}\} p_\gamma(x) Pr(I_k = 1 | \gamma_k = x) dx \quad \text{(B.10)}$$

In a SIMO system, maximum one data stream is sent to each user at a time, this gives $S = 1$ in (B.10). With a PF scheduler, a user with high instantaneous channel gain is more likely to be scheduled than another one with poor channel gain, as shown in (B.9). Inserting (B.2) and (B.9) into (B.10) leads to:

$$C_{SIMO,PF}(M, K, \alpha, \bar{\gamma}, C_{max}) = K \int_0^\infty min\{\log_2(1+\alpha x), C_{max}\} p_\gamma(x) Pr(I_k = 1 | \gamma_k = x) dx$$

$$= \frac{K}{\bar{\gamma}^M (M-1)!} \int_0^\infty min\{log_2(1+\alpha x), C_{max}\} x^{M-1} e^{-x/\bar{\gamma}} \left(1 - e^{-x/\bar{\gamma}} \sum_{m=0}^{M-1} \frac{(x/\bar{\gamma})^m}{m!}\right)^{K-1} dx \quad \text{(B.11)}$$

$$= \frac{K}{\bar{\gamma}^M (M-1)!} \int_0^\infty min\{log_2(1+\alpha x), C_{max}\} x^{M-1} e^{-x/\bar{\gamma}} \sum_{k=0}^{K-1} \binom{K-1}{k} (-1)^k e^{-kx/\bar{\gamma}} \left(\sum_{m=0}^{M-1} \frac{(x/\bar{\gamma})^m}{m!}\right)^k dx$$

After the expansion of the power of the summation, and some algebra, the above equation becomes:

$$C_{SIMO,PF}(M, K, \alpha, \bar{\gamma}, C_{max}) = \frac{K}{\bar{\gamma}^M (M-1)!} \sum_{k=0}^{K-1} (-1)^k \binom{K-1}{k} \sum_{n_0=0}^{k} \cdots \sum_{n_{M-2}=0}^{n_{M-3}} \\ \frac{\prod_{m=0}^{M-1} \binom{n_{m-1}}{n_m} (1/m!)^{n_{m-1}-n_m}}{\bar{\gamma}^{\sum_{m=1}^{M-1} m(n_{m-1}-n_m)}} \mathbb{F}\left(\alpha, \frac{k+1}{\bar{\gamma}}, M-1+\sum_{m=1}^{M-1} m(n_{m-1}-n_m), C_{max}\right) \quad \text{(B.12)}$$

with $n_{-1} = k$, and $n_{M-1} = 0$. The derivation of the function $\mathbb{F}(\alpha, \lambda, n, C_{max})$ is presented in the end of the appendix, and its expression is shown in (B.31). This function is also used for the expression of the other investigated multi-antenna schemes and packet schedulers. Equation (B.12) is applicable only when $M \geq 2$, the expression for a SISO system can be obtained by setting $M = 1$ in the MIMO capacity expression that is presented in the following.

In a MIMO system with ZF precoding, M data streams are sent in each time slot. According to the PDF distribution of (B.4) and the scheduling probability (B.9),

the system capacity with PF scheduling is:

$$C_{MIMO,PF}(M,K,\alpha,\bar{\gamma},C_{max}) = KM\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}p_\gamma(x)Pr(I_k=1|\gamma_k=x)dx$$

$$= KM^2/\bar{\gamma}\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}e^{-xM/\bar{\gamma}}\left(1-e^{-xM/\bar{\gamma}}\right)^{K-1}dx$$

$$= KM^2/\bar{\gamma}\sum_{k=0}^{K-1}(-1)^k\binom{K-1}{k}\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}e^{-xM(1+k)/\bar{\gamma}}dx$$

$$= KM^2/\bar{\gamma}\sum_{k=0}^{K-1}(-1)^k\binom{K-1}{k}\mathbb{F}(\alpha, M(1+k)/\bar{\gamma}, 0, C_{max}) \quad \text{(B.13)}$$

The performance with an RR scheduler can be derived in a similar manner as with PF. Alternatively, it can also be expressed using the derived formulas for PF. This is possible because the RR scheduler cannot exploit the multi-user diversity. Therefore, its average spectral efficiency is insensitive to the number of users in a system, and is identical to the performance of PF in a single-user case ($K=1$):

$$C_{SIMO,RR}(M,\alpha,\bar{\gamma},C_{max}) = C_{SIMO,PF}(M,1,\alpha,\bar{\gamma},C_{max})$$
$$= \frac{1}{\bar{\gamma}^M(M-1)!}\mathbb{F}(\alpha,1/\bar{\gamma},M-1,C_{max}) \quad \text{(B.14)}$$

$$C_{MIMO,RR}(M,\alpha,\bar{\gamma},C_{max}) = C_{MIMO,PF}(M,1,\alpha,\bar{\gamma},C_{max})$$
$$= (M^2/\bar{\gamma})\mathbb{F}(\alpha,M/\bar{\gamma},0,C_{max}) \quad \text{(B.15)}$$

In a system with 2 transmit and receive antennas, according to the discussion in Section B.2, the capacity for rank-1 transmission equals:

$$C_{2\times2,rank-1}(K,\alpha,\bar{\gamma},C_{max}) = C_{SIMO}(4,K,\alpha,\bar{\gamma}/2,C_{max}) \quad \text{(B.16)}$$

Equation (B.16) is valid for both RR and PF schedulers, therefore the dependence of the packet scheduler is dropped in the expression. Rank adaptation then chooses the proper rank to maximize the system capacity.

B.4 Numerical Results

In Section B.3, the capacity expressions for various antenna configurations and packet schedulers have been derived. With these expressions, the system performance can be obtained and the multi-user gain can be calculated from the relative capacity of PF as compared to a single user system (or with RR). For the numerical evaluation, the SINR scaling factor of $\alpha = 0.5$ is used, which corresponds to 3 dB loss from the Shannon boundary. With 64-QAM and coding rate of 9/10 as the highest modulation and coding scheme, $C_{max} = 5.4$ bps/Hz. The output of the derived expressions are compared with Monte Carlo simulations for verification.

B.4.1 Verification of the Capacity Expressions

First, both estimated and simulated capacity for PF with different antenna configurations is obtained, as shown in Fig. B.2. In order to see how multi-user diversity affects the system performance, the number of users are varied from 1 to 20, at an average SINR of 5 dB. RR has the same performance as single-user PF, therefore RR is not explicitly plotted in the figure. From this figure it can be seen that there is an excellent agreement between the analytical predictions and the simulated results, confirming the validity of the derived expressions. It is also observed that the relative performance between MIMO-ZF precoding and SIMO-MRC is sensitive to the level of multi-user diversity. When the number of users is low, it is better to exploit the diversity gain offered by multi-antenna transmission. When the number of users is high, the PF scheduler can find a good user for each data stream, therefore MIMO multiplexing is preferable.

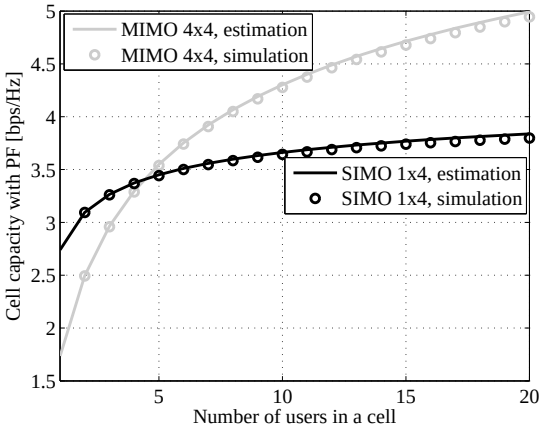


Fig. B.2: Estimated and simulated capacity with 1×4 and 4×4 antenna configurations and the PF scheduler. Performance is evaluated with different number of users, and an average SINR of 5 dB for each user.

Next the performance at different channel quality levels is analyzed. The number of users in a system is assumed to be 5, and the scheduler of both RR and PF are considered. The results are summarized in Fig. B.3. Similar as Fig. B.2, it is found that 4×4 MIMO with ZF precoding is not always beneficial as compared to a 1×4 SIMO system with an MRC receiver. In fact, at very low SINR values the performance of 4×4 MIMO with ZF precoding is even worse than 1×4 MRC. The reason is that, when the SINR is low, $\log_2(1+\gamma) \approx \log_2(e)\gamma$ and the capacity

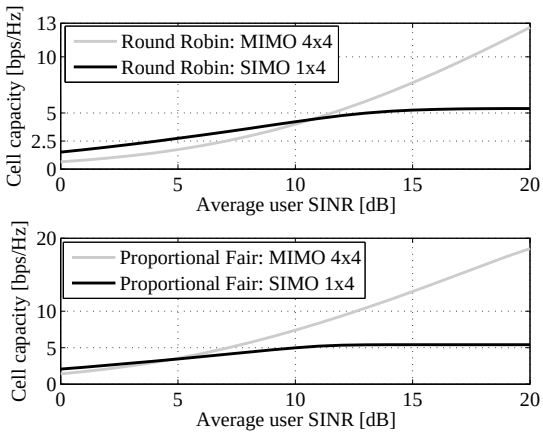


Fig. B.3: Capacity with 1×4 and 4×4 antenna configurations. Performance at different channel quality levels is shown, with 5 users in the system. The upper plot is for RR scheduler, the lower plot is for PF scheduler.

for the RR scheduler with $M \times M$ MIMO can be approximated as:

$$C_{MIMO,RR}(M, \alpha, \bar{\gamma}, C_{max}) \approx \log_2(e) M^2 / \bar{\gamma} \int_0^\infty \alpha x e^{-xM/\bar{\gamma}} dx$$

$$= \log_2(e) \alpha \bar{\gamma}, \qquad \bar{\gamma} \to 0^+ \tag{B.17}$$

This is the same as the performance of a SISO system. Meanwhile, MRC still has the array gain, leading to M times the performance of SISO:

$$C_{SIMO,RR}(M, \alpha, \bar{\gamma}, C_{max}) \approx \frac{\log_2(e)}{\bar{\gamma}^M (M-1)!} \int_0^\infty \alpha x e^{-x/\bar{\gamma}} x^{M-1} dx$$

$$= M \log_2(e) \alpha \bar{\gamma}, \qquad \bar{\gamma} \to 0^+ \tag{B.18}$$

B.4.2 Multi-user Gain and the Logarithmic Approximation

A PF scheduler exploits the multi-user diversity to achieve better performance than an RR scheduler. By comparing the bottom plot in Fig. B.3 with the top one, it can be seen that the multi-user gain is a decreasing function of user channel quality. This is because of the logarithmic behavior in (B.1). Moreover, with RR the performance of 4×4 MIMO-ZF is better than 1×4 MRC for SINR beyond 11 dB, whereas the corresponding switching point is less than 5 dB for PF. This is due to the fact that MIMO multiplexing reduces the average SINR per data

stream as compared to single stream transmission, and hence increases the multi-user gain. On the other hand, MIMO (or SIMO) diversity reduces the variation in channel quality and increases the average SINR; both have an effect of reducing the multi-user gain. Therefore, a channel aware packet scheduler favors MIMO multiplexing more than SIMO/MIMO diversity.

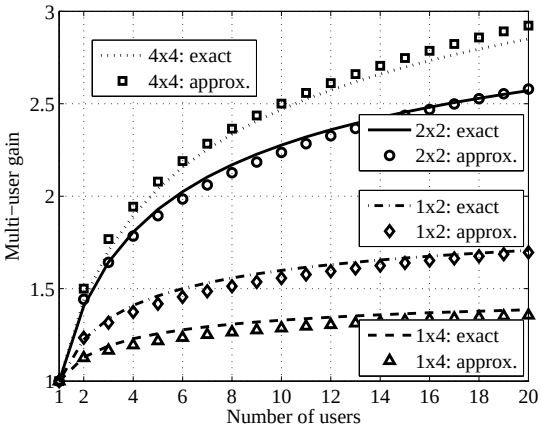


Fig. B.4: Multi-user gain with different antenna configurations, with average SINR of 5 dB.

The multi-user gain versus the number of users is plotted in Fig. B.4, for both SIMO and MIMO systems at an average SINR of 5 dB. It is observed that when the capacity limit C_{max} is not reached, the multi-user gain roughly follows a logarithm function of the number of users. By using curve fitting over a large variation of average channel levels and antenna configurations, with the target of minimizing the mean square error between the approximated and the original values, the following approximation function is obtained:

$$G_{MIMO}(M,K,\alpha,\bar{\gamma},C_{max}) \approx \begin{cases} 1, & K=1 \\ min\left\{G_{max}(K), 1.1+0.46\left(\frac{M}{\alpha\bar{\gamma}}\right)^{\log_{10}2}\ln(K)\right\}, & K>1 \end{cases} \quad \text{(B.19)}$$

$G_{max}(K)$ is the upper bound of multi-user gain, which is obtained when the SINR per data stream is so small that the mapping from SINR to spectral efficiency is linear. In this case, the average spectral efficiency with PF is

$$C_{MIMO,PF}(M,K,\alpha,\bar{\gamma},C_{max}) \approx \log_2(e)\alpha\bar{\gamma}K\sum_{k=0}^{K-1}\binom{K-1}{k}\frac{(-1)^k}{(1+k)^2}, \qquad \bar{\gamma}\to 0^+$$

Under the same condition, the capacity with RR is $\log_2(e)\alpha\bar{\gamma}$, as shown in (B.17).

Therefore,

$$G_{max}(K) = K \sum_{k=0}^{K-1} \binom{K-1}{k} \frac{(-1)^k}{(1+k)^2}$$

In fact, $G_{max}(K)$ is the same as the multi-user expression derived in [140], where the linear mapping from SINR to spectral efficiency is used. Equation (B.19) shows that the multi-user gain will increase with M, but decrease with $\alpha\bar{\gamma}$. This is due to the fact that the spectral efficiency varies relatively less in high SINR region than in low SINR region.

Different from MIMO with ZF precoding, the gain with MRC decreases with the number of receiving antennas. For very high values of M, or $\alpha\bar{\gamma}$, the variation in instantaneous spectral efficiency is small, and there will hardly be any multi-user gain. Similarly, with $1 \times M$ SIMO, the multi-user gain can be approximated in the logarithmic form:

$$\begin{aligned} & G_{SIMO}(M, K, \alpha, \bar{\gamma}, C_{max}) \qquad \text{(B.20)} \\ & \approx \begin{cases} 1, & K = 1 \\ max\left\{1, 1.14 - 0.02(M + \log_{10}(\alpha\bar{\gamma})) + \frac{0.46}{M(\alpha\bar{\gamma})^{\log_{10} 2}} \ln(K)\right\}, & K > 1 \end{cases} \end{aligned}$$

The approximated multi-user gain is also plotted in Fig. B.4, and it is close to the one obtained from simulation or the derived capacity expressions. It is worth noting that (B.19) and (B.20) are valid when the capacity limit is not reached. If the channel quality is so good that C_{max} is often reached, the multi-user gain will be significantly lower than shown here.

B.4.3 Performance of MIMO Rank Adaptation

The obtained results confirm that, when the channel quality is poor, it is better to use MIMO diversity to increase the signal strength rather than to transmit over multiple streams (the multiplexing gain). Therefore, MIMO rank adaptation is beneficial in the sense that it can adjust the number of transmitting data streams according to the channel condition. Using the derived equations of (B.13), (B.15) and (B.16), the performance of rank-1 and rank-2 transmissions in a 2×2 MIMO system can be calculated. The results in a system with 5 users are shown in Fig. B.5, where rank-1 achieves a better performance than rank-2 at low SINR values. In the high SINR region, rank-2 transmission overcomes the performance of rank-1. Also, PF favors the rank-2 transmission since it reduces the switching point for rank adaptation from 12.5 dB (RR) to 2.5 dB.

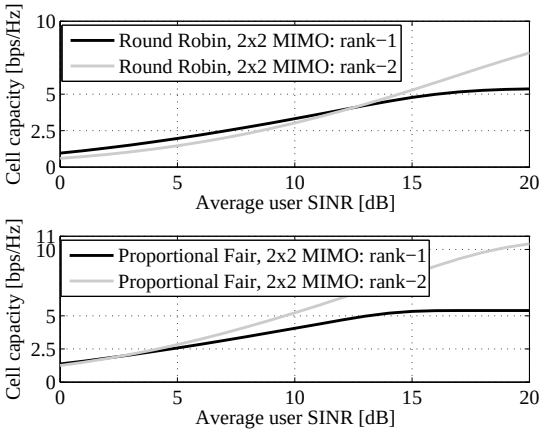


Fig. B.5: Capacity of a 2×2 MIMO system with different ranks. Performance at different channel quality levels is shown, with 5 users in a system. The upper plot is for RR scheduler, the lower plot is for PF scheduler.

B.5 Multi-user Gain in LTE Systems

The capacity expressions derived in Section B.3 are constricted by the assumption of having the same average channel quality for the users, so that PF is also fair in the sense of resource sharing. The accuracy of these expressions has been verified by simple simulations in Section B.4. It is observed from Section B.4 that the multi-user gain follows approximately a logarithmic function with the number of users in a cell. The aim of this section is to verify that the logarithmic behavior of the multi-user gain also exists in realistic LTE systems, where users are randomly distributed within the cell coverage area and have different average channel qualities. Other than the Macro-cell scenario that has been used in previous chapters, the Micro-cell scenario is also considered. Sectorization as in Macro-cell scenario is not used in the Micro-cell. The Micro-cell has a small coverage with inter site distance of 130 m and a different path loss model. As a consequence, the average channel quality with Micro-cell is better than with Macro-cell [147]. Details about the Micro-cell scenario can be found in [114].

With realistic assumptions as listed in Table A.3, the system performance and the multi-user gain are impacted due to the impairments. The multi-user gain in a Macro-cell in both average cell throughput and cell edge user throughput is presented in Fig. B.6. The performance is shown for bandwidths from 5 to 20 MHz, and clearly shows the independency of the multi-user gain from the system bandwidth (given that the bandwidth is wide enough). Despite the various system impairments, it is noticed that the multi-user gain follows a logarithmic function

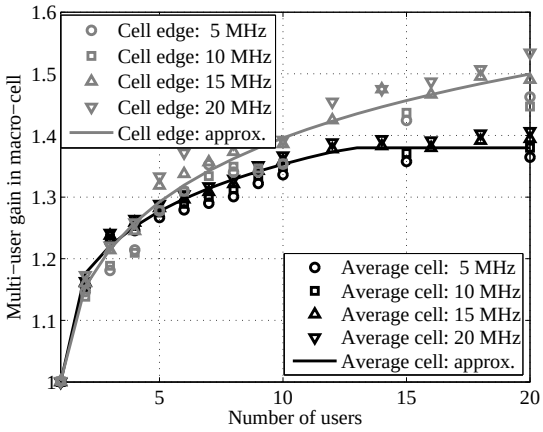


Fig. B.6: Multi-user gain in Macro-cell LTE with 1×2 SIMO. Simulation results for different bandwidth settings are approximated using logarithm functions.

of the number of users per cell. Furthermore, the gain saturates when the number of users is relatively high. This is mainly due to the granularity when performing scheduling and link adaptation. The curve-fitted expressions for the multi-user gain in the Macro-cell scenario are presented in (B.21) and (B.22) for both average cell and cell edge user throughput, respectively.

$$G_{macro,averge}(K) = \frac{C_{macro,average(K)}}{C_{macro,average(1)}} = \begin{cases} 1, & K = 1 \\ 0.11ln(K) + 1.10, & 1 < K \leq 13 \\ 1.38, & K > 13 \end{cases} \tag{B.21}$$

$$G_{macro,edge}(K) = \frac{KC_{macro,edge(K)}}{C_{macro,edge(1)}} = \begin{cases} 1, & K = 1 \\ 0.15ln(K) + 1.05, & K > 1 \end{cases} \tag{B.22}$$

Note that the multi-user gain in cell edge user throughput is scaled by the resource sharing per user. The above two equations are valid for a uniform distribution of users over the cell area, with K being the number of users per cell.

The corresponding performance in Micro-cell is presented in Fig. B.7, but for 10 MHz bandwidth configuration only. Similar as in the Macro-cell scenario, the multi-user gain can also be approximated using the logarithmic function:

$$G_{micro,average}(K) = \frac{C_{micro,average(K)}}{C_{micro,average(1)}} = \begin{cases} 1, & K = 1 \\ 0.05ln(K) + 1.06, & 1 < K \leq 11 \\ 1.18, & K > 11 \end{cases} \tag{B.23}$$

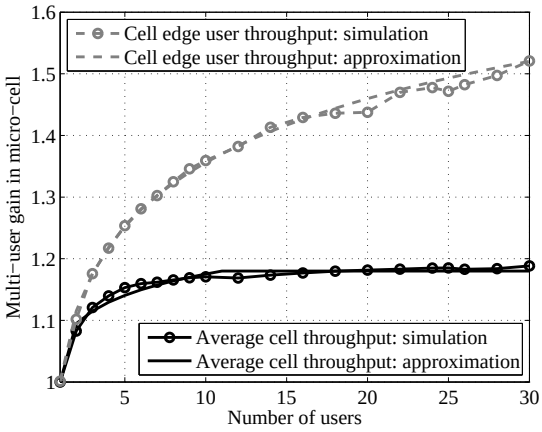


Fig. B.7: Multi-user gain in Micro-cell LTE with 1×2 SIMO. Simulation results are approximated using logarithm functions.

$$G_{micro,edge}(K) = \frac{KC_{micro,edge(K)}}{C_{micro,edge(1)}} = \begin{cases} 1, & K = 1 \\ 0.15ln(K) + 1.01, & K > 1 \end{cases} \tag{B.24}$$

B.6 Conclusion

In this appendix, the exact closed form expressions for the system capacity with channel blind (RR) and channel aware (PF) scheduling are derived, assuming uncorrelated Rayleigh fading channels and the same average channel quality among users. Several multi-antenna configurations are considered: the first one has a single transmit antenna and multiple receive antennas and uses MRC; the second one assumes the same number of transmit and receive antennas and uses ZF precoding. As a special case, the capacity expression for 2×2 MIMO with rank adaptation is also provided. Monte Carlo simulations are used to verify the derived expressions.

The derived capacity expressions can be used to analyze the system performance in various scenarios. Some representative system configurations are selected and their capacity is calculated. The obtained results confirm that in a multi-user system a channel aware packet scheduler can exploit the multi-user diversity to improve the system performance. This multi-user diversity gain increases with the number of users in a system, but decreases with the average user channel quality. When MIMO transmission is used, the MIMO diversity gain and multiplexing gain can be traded off with each other to obtain the best performance. The selection of the proper MIMO rank is affected by both the average channel quality and the

level of multi-user diversity in the system.

Based on the obtained results, it is found that the multi-user gain can be approximated by a logarithmic function in the number of users. This finding is also valid in LTE systems with realistic assumptions, and the approximation expressions are obtained for a couple of cases of interest.

Derivation of the Function $\mathbb{F}(\cdot)$ for the Capacity Expressions

This section provides the derivation of the integration $\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}e^{-\lambda x}x^n dx$, which is used in the expressions of the system capacity in Section B.3. Starting first with a variable substitution $y = 1 + \alpha x$:

$$\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}e^{-\lambda x}x^n dx$$

$$= \log_2 e \int_0^\infty min\{ln(1+\alpha x), C_{max}ln(2)\}e^{-\lambda x}x^n dx \quad \text{(B.25)}$$

$$= \frac{\log_2 e}{\alpha} \int_1^\infty min\{ln(y), C_{max}ln(2)\}e^{-\lambda(y-1)/\alpha}(y-1)^n\alpha^{-n}dy \quad \text{(B.26)}$$

By using the binomial expansion and interchanging between the integration and finite sum, (B.25) becomes:

$$\int_0^\infty min\{\log_2(1+\alpha x), C_{max}\}e^{-\lambda x}x^n dx$$

$$= \frac{e^{\lambda/\alpha}\log_2 e}{\alpha^{n+1}} \sum_{t=0}^{n} \binom{n}{t}(-1)^t \int_1^\infty min\{ln(y), C_{max}ln(2)\}e^{-\lambda y/\alpha}y^{n-t}dy$$

$$= \frac{e^{\lambda/\alpha}\log_2 e}{\alpha^{n+1}} \sum_{t=0}^{n} \binom{n}{t}(-1)^t \int_1^{2^{C_{max}}} ln(y)e^{-\lambda y/\alpha}y^{n-t}dy$$

$$+\frac{C_{max}e^{\lambda/\alpha}}{\alpha^{n+1}} \sum_{t=0}^{n} \binom{n}{t}(-1)^t \int_{2^{C_{max}}}^\infty e^{-\lambda y/\alpha}y^{n-t}dy$$

$$= \Psi + \Phi \quad \text{(B.27)}$$

where Ψ and Φ are the first and the second component of (B.27). Let $f(y) = ln(y)$, and $g(y) = \int e^{-\lambda y/\alpha} y^{n-t} dy = -e^{-\lambda y/\alpha} \sum_{h=0}^{n-t} \frac{(n-t)!}{h!} \left(\frac{\alpha}{\lambda}\right)^{n-t-h+1} y^h$. According to the rule of integration by parts, i.e., $\int_a^b f(y)g'(y)dy = f(y)g(y)|_b^a - \int_a^b f'(y)g(y)dy$, Ψ is re-written as

$$\begin{aligned}
\Psi &= -\frac{C_{max} e^{\lambda/\alpha - 2^{C_{max}}\lambda/\alpha}}{\lambda^{n+1}} \sum_{t=0}^{n} \binom{n}{t} (-1)^t \sum_{h=0}^{n-t} \frac{(n-t)!}{h!} \left(\frac{\lambda}{\alpha}\right)^{t+h} 2^{hC_{max}} \\
&+ \frac{e^{\lambda/\alpha} \log_2 e}{\lambda^{n+1}} \sum_{t=0}^{n} \binom{n}{t} (-1)^t \int_1^{2^{C_{max}}} e^{-\lambda y/\alpha} \sum_{h=0}^{n-t} \frac{(n-t)!}{h!} \left(\frac{\lambda}{\alpha}\right)^{t+h} y^{h-1} dy \\
&= \Psi_1 + \Psi_2 \qquad \text{(B.28)}
\end{aligned}$$

In (B.28), Ψ_1 and Ψ_2 are the first and the second component of Ψ. Interchanging integration and finite sum for Ψ_2 gives:

$$\begin{aligned}
\Psi 2 &= \frac{e^{\lambda/\alpha} \log_2 e}{\lambda^{n+1}} \sum_{t=0}^{n} \binom{n}{t} (-1)^t \sum_{h=0}^{n-t} \frac{(n-t)!}{h!} \left(\frac{\lambda}{\alpha}\right)^{t+h} \int_1^{2^{C_{max}}} e^{-\lambda y/\alpha} y^{h-1} dy \\
&= \frac{e^{\lambda/\alpha} \log_2 e}{\lambda^{n+1}} \sum_{t=0}^{n} \frac{n!}{t!} \left(-\frac{\lambda}{\alpha}\right)^t \sum_{t=0}^{n-t} \left(\Gamma(h, \lambda/\alpha) - \Gamma(h, 2^{C_{max}}\lambda/\alpha)\right)/h! \qquad \text{(B.29)}
\end{aligned}$$

where $\Gamma(s,x) = \int_x^\infty t^{s-1} e^{-t} dt$ is the incomplete gamma function with $s = 0, 1, 2, \ldots$. $\Gamma(0,x) = \int_x^\infty e^{-t}/t dt = Ei(x)$ is the exponential integral function, and $\Gamma(1,x) = e^{-t}$. The second component in (B.27) is:

$$\begin{aligned}
\Phi &= \frac{C_{max} e^{\lambda/\alpha}}{\alpha^{n+1}} \sum_{t=0}^{n} \binom{n}{t} (-1)^t \int_{2^{C_{max}}}^\infty e^{-\lambda y/\alpha} y^{n-t} dy \\
&= \frac{C_{max} e^{\lambda/\alpha}}{\alpha^{n+1}} \sum_{t=0}^{n} \binom{n}{t} \left(-\frac{\lambda}{\alpha}\right)^t (n-t)! e^{-2^{C}_{max}\lambda/\alpha} \sum_{h=0}^{n-t} \frac{(2^{C_{max}}\lambda/\alpha)^h}{h!} \\
&= \frac{C_{max} e^{\lambda/\alpha - 2^{C_{max}}\lambda/\alpha}}{\lambda^{n+1}} \sum_{t=0}^{n} \binom{n}{t} (-1)^t \sum_{h=0}^{n-t} \frac{(n-t)!}{h!} \left(\frac{\lambda}{\alpha}\right)^{t+h} 2^{hC_{max}} \qquad \text{(B.30)}
\end{aligned}$$

From (B.28) and (B.30) it can be seen that Φ and Ψ_1 cancel each other, therefore, the integration of (B.25) is:

$$\begin{aligned}
&\int_0^{\infty} min\{\log_2(1+\alpha x), C_{max}\} e^{-\lambda x} x^n dx \\
&= \Phi_1 + \Phi_2 + \Psi \\
&= \frac{e^{\lambda/\alpha} \log_2 e}{\lambda^{n+1}} \sum_{t=0}^{n} \frac{n!}{t!} \left(-\frac{\lambda}{\alpha}\right)^t \sum_{t=0}^{n-t} \left(\Gamma(h, \lambda/\alpha) - \Gamma(h, 2^{C_{max}}\lambda/\alpha)\right)/h! \\
&= \mathbb{F}(\alpha, \lambda, n, C_{max})
\end{aligned} \tag{B.31}$$

If not considering the maximum capacity limit, $\Gamma(s, \infty) = 0$ and (B.31) degenerates to:

$$\begin{aligned}
&\int_0^{\infty} \log_2(1+\alpha x) e^{-\lambda x} x^n dx \\
&= \frac{e^{\lambda/\alpha} \log_2 e}{\lambda^{n+1}} \sum_{t=0}^{n} \frac{n!}{t!} \left(-\frac{\lambda}{\alpha}\right)^t \sum_{t=0}^{n-t} \frac{(n-t)!}{h!} \Gamma(h, \lambda/\alpha) = \mathbb{F}(\alpha, \lambda, n)
\end{aligned} \tag{B.32}$$

In the case of $M \times M$ MIMO and ZF precoding, n equals 0 in (B.25), which gives:

$$\begin{aligned}
&\int_0^{\infty} min\{\log_2(1+\alpha x), C_{max}\} e^{-\lambda x} x^n dx \\
&= \frac{e^{\lambda/\alpha} \log_2 e}{\lambda} \left(Ei(\lambda/\alpha) - Ei(2^{C_{max}}\lambda/\alpha)\right) = \mathbb{F}(\alpha, \lambda, 0, C_{max})
\end{aligned} \tag{B.33}$$

Appendix C

Theoretical Analysis of the Finite Buffer Transmission

In Chapter 4, the multi-Component Carrier (CC) system performance has been analyzed, mainly based on simulations. Both the full buffer and finite buffer traffic models are used and the full buffer results are verified by theoretical estimations. This appendix provides the theoretical analysis of the finite buffer transmission. The purpose is to verify the accuracy of the finite buffer simulation results.

The developed analytical model is valid for both single-CC and multi-CC systems. For the multi-CC systems, the analysis considers the case when all users are LTE-Advanced (assigned on all CCs), or all are Rel'8, which relies on carrier load balancing, e.g., Least Load (LL) or Mobile Hashing (MH), to select one CC per user.

C.1 The Birth-Death Model

The arrival, or departure, of users in a network is usually modeled as a birth-death process, as shown in Fig. C.1. The birth-death process is a special case of a continuous time Markov process, where the states represent the current number of active users, and the transitions are between neighboring states. The 'birth' is the transition towards increasing the active number of users by 1, and a 'death'

is the transition towards decreasing the number of active users by 1 [148]. The following notations are used for the considered traffic model:

S_k	System state with k users being served
λ_k	Arrival rate in state k, in users per second
μ_k	Service rate in state k; in users per second (the average service time is μ_k^{-1})
A	Maximum number of users each cell can serve

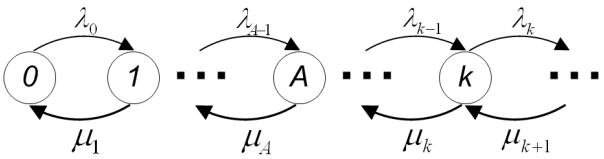


Fig. C.1: The birth-death process in queueing modeling. λ_k is the arrival rate going from state k to state $k+1$ (a 'birth'); μ_k is the departure rate of going back to state $k-1$ (a 'death').

The admission control in LTE(-Advanced) is assumed to limit the number of users per cell to maximum A users. This leads to the following arrival rate:

$$\lambda_k = \begin{cases} \lambda, & 0 \leq k < A \\ 0, & k \geq A \end{cases} \tag{C.1}$$

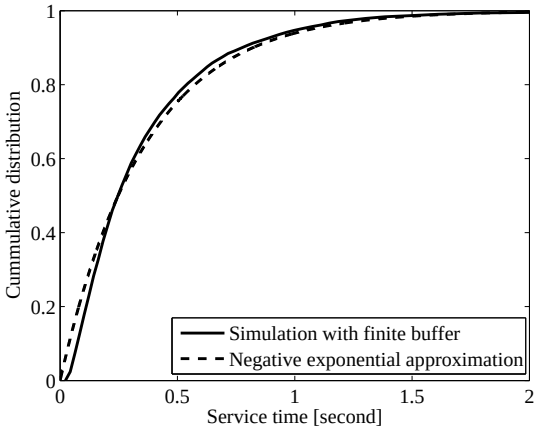


Fig. C.2: Service time: simulation result and its approximation.

The service time (time required for a user to finish its transmission) is usually assumed to follow a negative exponential distribution in voice transmission [149].

This assumption is roughly valid also for the data traffic, where users take different time to download a finite buffer due to channel quality variations. In Fig. C.2, the Cumulative Distribution Function (CDF) for the service time with arrival rate of 20/second/cell, and buffer size of 2 Mbits is shown. The LTE-Advanced system as described in Table A.3 is used, with independent scheduling per CC. According to simulation, the average service time is 0.36 second. The approximation is therefore performed using a negative exponential distribution, with CDF of $F(x) = 1-e^{-\lambda x}$, and $\lambda = 1/0.36 = 2.8$. This figure shows that in general the approximation is accurate, with small deviations for users that have very short service time.

With the exponential assumption for the service time, the probability of the system being at state S_k is given by [54]:

$$P_k = \begin{cases} P_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}}, & 1 \leq k \leq A \\ 0, & k > A \end{cases} \tag{C.2}$$

where P_0 is the probability of being in state S_0, and

$$\sum_{k=0}^{A} P_k = 1 \tag{C.3}$$

Inserting (C.2) into (C.3), the following equation is obtained:

$$P_0 = \left(1 + \sum_{k=1}^{A} \prod_{i=1}^{k} (\lambda_i / \mu_{i+1}) \right)^{-1} \tag{C.4}$$

C.2 Analysis When Each User Can Access the Whole Bandwidth

This section provides the analytical model for the case when each user can access the whole bandwidth, i.e., a single-CC system, or a multi-CC system with only LTE-Advanced users.

C.2.1 Analytical Estimations

By using channel aware packet scheduling, the multi-user diversity can be exploited to improve the system performance. Therefore, the service rate at state S_k (k users served) can be expressed by using the throughput and the buffer size, as

$$\mu_k = \begin{cases} \frac{C}{F} \frac{G(k)}{G(\infty)}, & 0 \leq k < A \\ 0, & k \geq A \end{cases} \tag{C.5}$$

where C is the average cell throughput, obtained when the cell load is high enough such that the multi-user gain saturates; F is the buffer size for the user packet transmission; $G(\cdot)$ is a logarithmic function to approximate the multi-user gain, as presented in (B.21).

Inserting (C.1) and (C.5) into (C.2), the probability of the system being in state S_k can be expressed as

$$P_k = \begin{cases} P_0 \prod_{i=0}^{k-1} \frac{\lambda_i}{\mu_{i+1}} = P_0 (\lambda F/C)^k \prod_{i=1}^{k} G(\infty)/G(i), & 1 \le k \le A \\ 0, & k > A \end{cases} \tag{C.6}$$

Equation (C.6) is a function of the arrival rate λ, the buffer size F, the capacity limit C, the multi-user gain $G(\cdot)$ and the admission control limit A. Once these assumptions are known, an estimation of the system being in each state can be obtained.

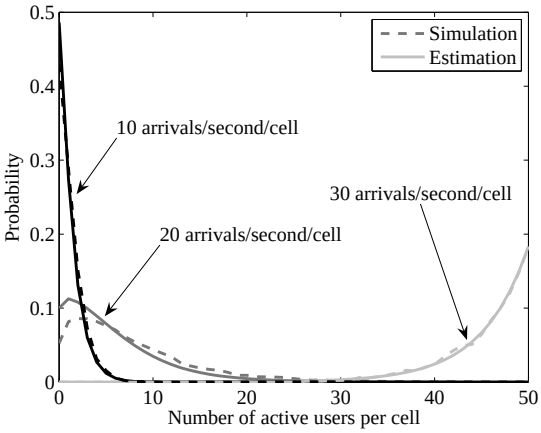


Fig. C.3: Probability of a cell being in different states of a Markov chain.

It has been shown in [108] that in a system with assumptions listed in Table A.3 and using the finite buffer model, the capacity in a 40 MHz system is 49 Mbps, i.e., $C = 49$ in (C.6). With maximum 50 admitted users ($A = 50$), the probability at different arrival rates are obtained both using (C.6) and from simulation, which is plotted in Fig. C.3. As can be seen from this figure, there is a good agreement between the theoretical estimation and the simulation result. As the arrival rate increases, the system is more likely to be in a state with more active users. If the arrival rate is larger than the cell can accommodate, some users cannot get enough resources, and they will remain in the system; in this case, the number of users will increase over time to A, and thereafter remain at that level.

The average number of active users per cell can also be estimated using (C.6), as

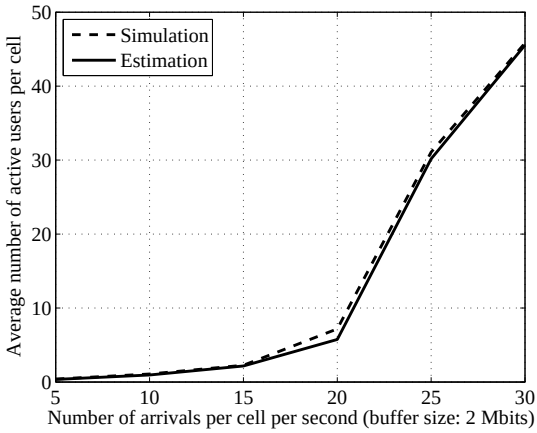


Fig. C.4: Average number of active users at different arrival rates.

$\tilde{K} = \sum_{k=1}^{A} kP_k$. This is compared with simulation results in Fig. C.4, and again, with a good match between the two.

C.2.2 Comparison between Single and Multi-CC Systems

The analysis before is valid for both the single-CC case and the multi-CC case when only the LTE-Advanced users exist. A comparison between these two systems is given in this section. For a fair comparison, the buffer size and the admission control limit is kept the same, but the arrival rate is proportional to the system bandwidth, i.e., the 4-CC system is evaluated with 4 times the arrival rates of the single-CC system.

According to (C.6), the probability of one CC having k users is the same for the two systems, therefore, the user throughput should be proportional to the system bandwidth. It is also quite intuitive, that more spectrum should bring more throughput. This is witnessed in Fig. C.5, where a 4-CC system achieves 4 times the average cell throughput and cell edge user throughput of a single CC system. Therefore, it verifies that the implementation of the bandwidth extension is correct.

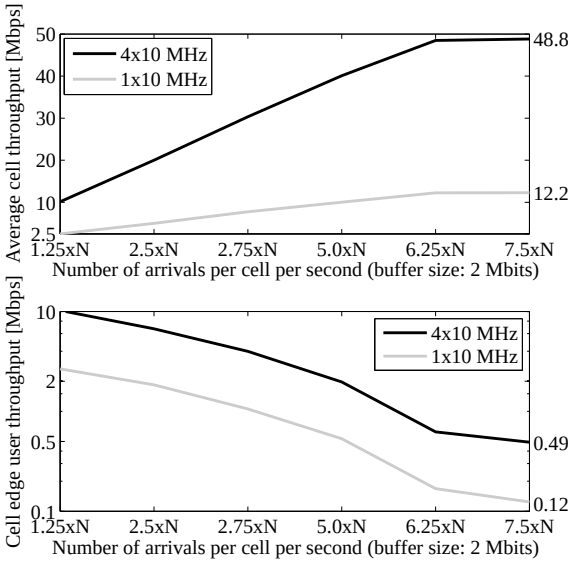


Fig. C.5: Performance comparison: single-CC versus 4-CC.

C.3 Analysis in Multi-CC Systems With Load Balancing

With LTE-Rel'8 users exist in a multi-CC system, each Rel'8 user has access to only one CC, which is decided by the CC assignment function. In Chapter 4, two load balancing algorithms have been considered to assign the CCs, i.e., the LL balancing and the MH balancing. One can refer to Section 4.2.1 on how they work. Fig. C.6 shows the normalized cell edge user throughput for the two balancing methods, with respect to assigning each user on all CCs. For MH balancing, the probability of the system to reach the admission control limit (P_{50}) is also plotted, in order to decide whether or not the system is capacity-limited.

For MH balancing, when arrival rate is very low and the admission control limit is not reached, the arrival on each CC also follows the Poisson distribution, with $1/N$ the cell arrival rate (N is the number of CCs). It is equivalent to the single-CC system analyzed before, and should achieve $1/N$ the performance with LTE-Advanced users. However, when arrival rate is high, the arrival on one CC might be blocked because the other CCs are heavily loaded. In this situation, the arrival per CC is not Poisson arrival. This confirms the simulation results, that for the arrival rate below 20 the system with MH balancing has enough capacity to accommodate the user arrival ($P_{50} = 0$), and it achieves around 25% the performance when each

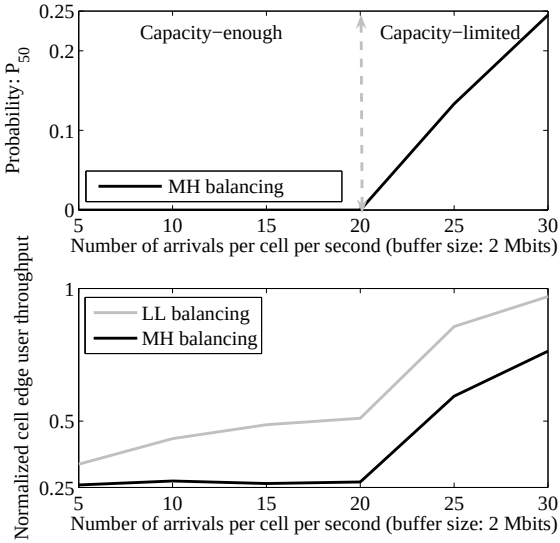


Fig. C.6: Performance comparison: multi-CC system with different load balancing methods.

user can access all CCs.

For LL balancing, the arrived user will be assigned to the CC that has the lowest load. It tries to balance the load across the CCs and maintains better fairness than MH. As a result, the cell edge user throughput is higher than MH. At very high load, the system stabilizes with A/N users per CC and each user gets the same average amount of resources as the case with LTE-Advanced users. Therefore, LL balancing has the same performance as assigning each user on all CCs.

The above analysis also verifies the simulation results in the other chapters of the thesis, e.g., in Fig. 4.10 when independent scheduling is used and all users are LTE-Advanced, it achieves 4 times the performance of MH balancing; in Fig. 5.5 when BCL is used to select one CC for each user (LL balancing), the performance is higher than 25% as compared to the case when each user has access to all CCs.

C.4 Conclusion

This appendix provides an analytical model for the finite buffer transmission. Given the cell capacity, the arrival rate and the buffer size, it is able to estimate

the probability of a system having different number of active users, and the average number of active users per cell.

The estimation matches the simulation, and therefore confirms that the implementation of the finite buffer traffic with Poisson arrival is correct. Using this model, the relative behavior of the single and multi-CC system can also be estimated, which is the same as obtained from simulation, indicating the extension from single to multi-CC system is correct. Furthermore, the performance of the LL and MH balancing methods at different load conditions is also analyzed, which verifies the accuracy of the simulation results.

Appendix D

Further Improvement of the Wide Area Network Performance

The existence of multiple Component Carrier (CC)s gives challenges and opportunities for the Radio Resource Management (RRM) of the Long Term Evolution (LTE)-Advanced system. The baseline performance of the multi-CC LTE-Advanced systems has been shown in Chapter 4, and the feedback overhead reduction for such systems has been analyzed in Chapter 5. Nevertheless, there are many other techniques that can potentially improve the system performance. Some of these techniques are multi-CC dependent, whereas others are in general applicable to both single and multi-CC systems.

In this appendix, some representative techniques are selected and their capability in improving the system performance will be demonstrated. Section D.1 discusses the network utility maximization as well as the trade-off between average cell throughput and cell edge user throughput using Generalized Proportional Fair (GPF). Section D.2 deals with the uplink feedback reduction. A load adaptive average best-M report method is investigated to improve the performance over fixed best-M. Section D.3 proposes a method to achieve different frequency reuse patterns in a multi-CC system by cell-level CC assignment, and the performance is evaluated for various reuse cases. In Section D.4, a scenario with mixed velocity users is considered, and a user velocity dependent CC assignment is investigated. Finally, conclusions are drawn in Section D.5.

D.1 Generalized Proportional Fair Scheduling

D.1.1 Introduction of the GPF Scheduling

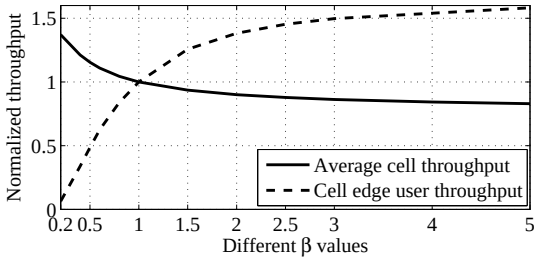
In Section 4.2.2 it has been proved that the cross-CC Proportional Fair (PF) scheduling maximizes the log-measure utility of a cell, giving the condition that each user's throughput is averaged over a time window which is long enough. Also, it is well-known that the traditional independent PF maximized such utility in a single CC system under the same condition. In [96], a GPF algorithm has been proposed to trade off between average user (or cell) throughput and cell edge user throughput, by weighting the instantaneous and past averaged throughput differently. However, the utility maximization of PF is based on the condition of enough long average window for the user throughput. This is easily satisfied with the full buffer model, where users are kept alive all the time. For the finite buffer model where each user has a certain buffer to transmit/receive, the user alive time may be rather short when the channel quality is good, or the buffer size is small. Therefore, the condition for the utility maximization may not be satisfied, and it is questionable whether or not GPF can keep its original goal of trading off between the two throughput indicators. This section is devoted to investigate the behavior of GPF with different traffic models, as well as to justify its usefulness.

GPF employs a scheduling metric of $r(k,n,m,t)^{\phi}/R(k,n,t)^{\theta}$, where $r(k,n,m,t)$ is the instantaneous throughput for user k in the n^{th} CC on the m^{th} resource element at time t, and $R(k,n,t)$ is the average throughput for user k in the n^{th} CC until time t [96]. It can easily be extended to cross-CC scheduling by replacing the denominator with the user throughput aggregated over the CCs, as presented in Section 4.2.2. Furthermore, maximizing this metric is equivalent to maximizing the metric $r(k,n,m,t)/R(k,n,t)^{\theta/\phi} = r(k,n,m,t)/R(k,n,t)^{\beta}$, with $\beta = \theta/\phi$. Therefore, the cross-CC GPF scheduling metric can be written as: $r(k,n,m,t)/R(k,t)^{\beta}$.

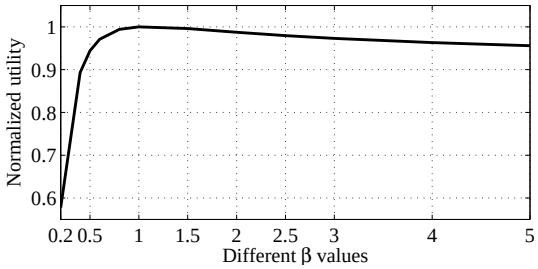
GPF is the same as normal PF when $\beta = 1$, and it maximizes the utility. When $\beta < 1$, it favors the users with good channel quality similar to the maximum rate scheduler; $\beta > 1$ favors the poor users towards an equal throughput scheduler [96].

D.1.2 Performance of GPF Scheduling

The same LTE-Advanced downlink transmission system as described in Table A.3 is used for the performance evaluation, and the full buffer model with 10 users per cell is used at first. The user throughput is initialized based on the channel quality and the expected spectrum that a user would occupy. It is time-averaged with an average window of 400 ms. All users are assumed to be LTE-Advanced, which



(a) Normalized throughput

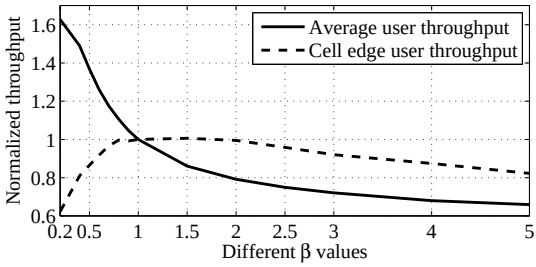


(b) Normalized utility

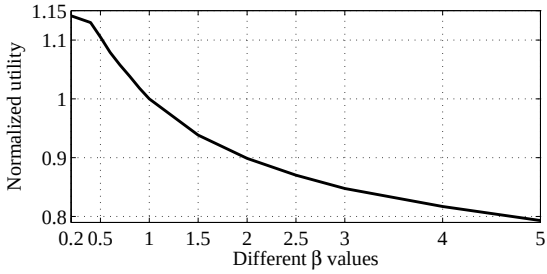
Fig. D.1: Normalized performance of GPF with different β values with respect to PF. Full buffer model is used with 10 users per cell.

can access all the 4 aggregated CCs. The performance is summarized in Fig. D.1, where Fig. D.1(a) shows the normalized average user (or cell) throughput and cell edge user throughput with respect to the normal PF scheduling (GPF with $\beta = 1$), and Fig. D.1(b) presents the network utility, which is the summation of the log-measure throughput for users within a cell, according to (4.7). The utility is also normalized with respect to what a normal PF can achieve. As can be seen from this figure, a small value of β improves the average user/cell throughput because more resources are assigned to good users. Meanwhile, the cell edge user throughput decreases because the poor users get less transmission resources. On the opposite, a high value of β improves the cell edge user throughput at the cost of poor average user/cell throughput. This confirms the existence of the trade-off between the two throughput indicators. Also, Fig. D.1(b) indicates that GPF with $\beta = 1$ maximizes the network utility. The presented results for the 4-CC system is equivalent to a single-CC system, except for a constant shift in the utility value (due to a scaling factor of 4 in the throughput).

With the finite buffer model, the number of users in a cell changes with time, hence the same equation used for the full buffer model cannot be used here. Instead, the utility is calculated as the summation of the log-measure throughput for all the served users during a certain time period. This time period should be long enough to capture the variation in the number of users per cell. The simulations for different β values are carried out at an arrival rate of 20 per cell per second, for a duration of 150 seconds, which is also the time window for the utility calculation. The buffer size is 2 Mbits per user, which is the same as in previous chapters.



(a) Normalized throughput



(b) Normalized utility

Fig. D.2: Normalized performance of GPF with different β values with respect to PF. Finite buffer model is used with 20 arrivals per cell per second.

Fig. D.2 shows the performance of GPF with different β values, normalized at $\beta = 1$. As indicated in Fig. D.2(a), the normalized average user throughput is similar to the full buffer case, which decreases with β. However, the cell edge user throughput remains roughly constant for $0.8 \leq \beta \leq 1.8$, and any other value of β causes a loss. Therefore, when the finite buffer model is used, the cell edge user throughput cannot be improved by sacrificing the average user throughput. This is different from the conclusion drawn from the full buffer model. Looking at the utility plot in Fig. D.2(b), it can be seen that the utility constantly decreases with

β and the maximization of the utility does not happen when $\beta = 1$.

Fig. D.2(a) also reveals that $\beta = 0.8$ can improve the average user throughput without penalizing the cell edge user throughput. $\beta > 1$ should not be used because it reduces the average user throughput with no gain in the cell edge throughput; the network utility is also lower than with a small β value. The different behavior of GPF with the traffic model is because: when the finite buffer model is used, $\beta < 1$ assigns more resources to the users with good channel quality. These users finish their transmissions in a short duration, and release the resources to the poor users after their departure. This differs from the full buffer case, where all users are kept alive, and the good users are constantly favored by a small value of β.

In Fig. D.3, the CDF for GPF with $\beta = 0.8$ and $\beta = 1.0$ can be seen. It clearly shows that $\beta = 0.8$ improves the throughput of the good users with no loss of the poor users' performance as compared to $\beta = 1.0$.

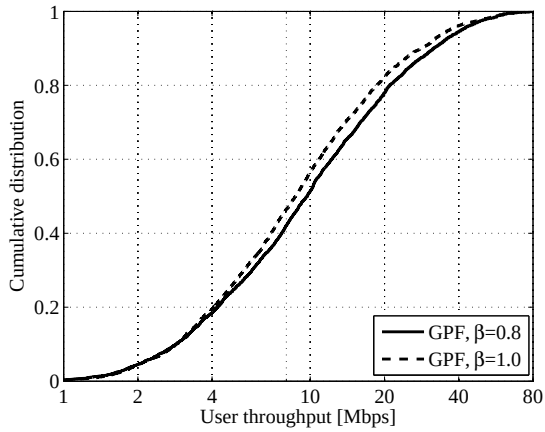


Fig. D.3: CDF of the user throughput for GPF with different β values. Finite buffer model is used with 20 arrivals per cell per second.

The analyses before are made with 20 arrivals per cell per second, which corresponds to an offered load of 40 Mbps. It is also worth taking a look at the behavior of GPF at different load conditions. For this purpose, performance is evaluated at different arrival rates from 5 to 20, and the results are summarized in Fig. D.4. For the cell edge user throughput, it can be seen that $\beta = 0.8$ achieves almost the same performance as $\beta = 1$ (PF) in all load conditions. As regards the average user throughput, the relative behavior of $\beta = 0.8$ over $\beta = 1.0$ increases with the offered load. This is expected because when the arrival rate is very low, there will most likely be 0 or 1 active users per cell. If this is the case, the active user will be assigned the whole bandwidth despite the scheduling algorithms. As the

arrival rate increases, more and more users will simultaneously be active in a cell, and hence the effect of the packet scheduling becomes obvious. The resource utilization versus the cell load (user arrival rate) can be seen in Fig. D.5. Clearly, it increases with the cell load and is not very sensitive to the value of β. With an arrival rate of 20 per second, the resource is almost fully utilized($\sim$95%), and the gain in average user throughput by using GPF with $\beta = 0.8$ is 10% over PF.

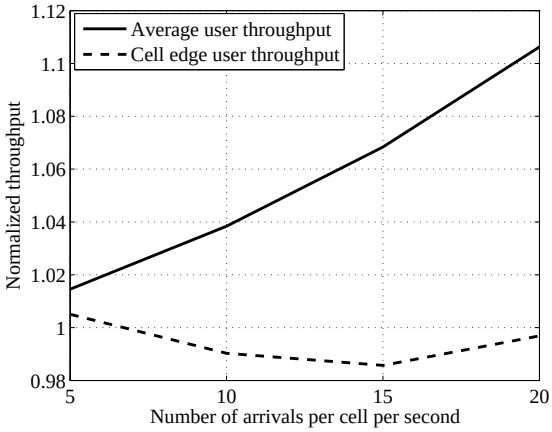


Fig. D.4: Normalized performance of GPF with $\beta = 0.8$ with respect to PF. Finite buffer model is used with various arrival rates.

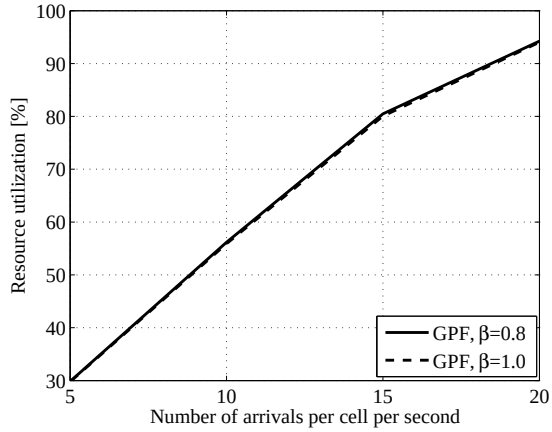


Fig. D.5: Resource utilization versus the cell load (arrival rate). Results are shown for GPF with different β values.

D.1.3 Summary on GPF With Different Traffic Models

In this section the performance of GPF scheduling, which tries to trade off between average user (or cell) throughput with the cell edge user throughput, is evaluated. The impact of traffic models on the packet scheduling is also studied. It is found that with the full buffer model, GPF maximizes the network utility when $\beta = 1$, and the trade-off exists between the average user/cell throughput and the cell edge user throughput. However, when the finite buffer model is used, it is more advantageous to use a low value of β (0.8) to favor the good users. It improves the average user throughput and maintains the same cell edge user throughput. The gain in average user throughput is sensitive to the cell load, which increases from 0% at very low load to ∼10% when the load approaches the cell capacity. GPF is a simple extension of the traditional PF scheduler, and offers more flexibility in the resource allocation, therefore is more preferable than PF for the resource sharing among best effort users.

D.2 Load Adaptive Average Best-M Report

D.2.1 Problem Analysis

The feasibility of using average best-M to reduce the CQI overhead as compared to full CQI report has been analyzed in Section 5.2, and it is observed that average best-M is a very efficient algorithm to reduce the feedback overhead while maintaining good throughput performance. The relative performance of average best-M as compared to full CQI report is affected by the value of M, as well as the cell load. However, according to [27], a fixed value of M is used for each bandwidth configuration, which is not fully optimized with respect to cell load. Changing the value of M has the following affect on the system performance:

1. As M increases from 0 to N_{CQI} (the number of CQI groups in a CC), the overhead will first increase, then decrease. The maximum overhead happens when $M \approx N_{CQI}/2$. In this thesis, 4 bits are used to quantize one CQI value, i.e., $Q_{CQI} = 4$. In a 10 MHz CC, $N_{CQI} = 17$, the overhead required by full CQI report is $N_{CQI} \times Q_{CQI} = 68$ bits. Using average best-M, the feedback overhead with different M values is calculated according to (5.1), and the normalized values (with respect to full CQI report) are plotted in Fig. D.6. Typically, M has the value of 4 or 5 [27], giving 72∼74% overhead reduction.

2. A high value of M increases the possibility that a user is scheduled on its best CQI groups.

3. As M increases, the accuracy of the best CQIs will get lower, because the best CQIs are encoded using only one offset value from the average CQI. For a very high M value, the best CQI report will approach the wideband CQI. Therefore it is not suggested to have M much larger than $N_{CQI}/2$.

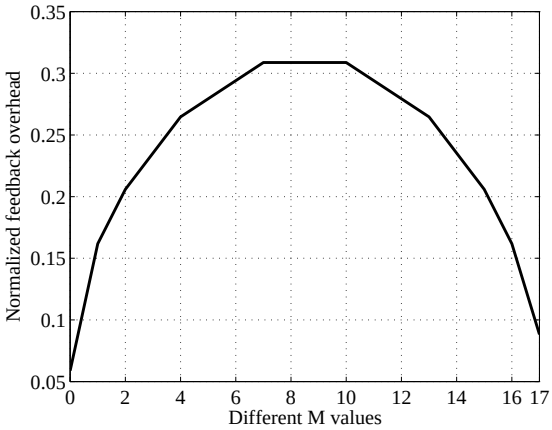


Fig. D.6: Feedback overhead for average best-M report with different M values, normalized with full CQI report.

While the first point gives information about the overhead reduction, the second and the third points affect the system throughput in opposite manner. E.g, a high value of M increases the possibility that one user is scheduled on its best CQIs (good), but decreases the accuracy of the best CQI values (bad). In an extreme case when $M = N_{CQI}$, the performance will be the same as having wideband CQI report only, because all CQI groups are considered as "best" ones. On the other hand, $M = 0$ is able to provide the best performance if there is only a single user per cell. This is true because the single user will be scheduled and encoded across the whole bandwidth, hence only the wideband CQI value is used, even if the feedback contains more detailed information (when $0 < M < N_{CQI}$).

Based on the above discussions, the performance of average best-M does not always increase (or decrease) as M goes high. In order to maximize the throughput performance while maintaining low CQI overhead, the value of M should be properly set according to the load condition. The adaptive best-M has been studied in [150, 151]. However, a best-M scheme that reports the individual CQI values of the selected CQI groups is used in these studies. The individual best-M differs from the average best-M in the sense that the overhead increases linearly with M; the throughput is also a strictly increasing function of M. To optimize the performance of average best-M, the adaptation of M addressing its above listed characteristics is needed. This is dealt with in this section.

D.2.2 The Proposed Load Adaptive Method

In order to optimize the performance of average best-M at different cell loads, an estimation of the performance for different M values is given at first. Consider a system with average best-M, each CQI group has probability M/N_{CQI} to be considered as "best" by one user. With K users, the probability for a CQI group to be considered as "best" by K' users is $P_{K'} = \binom{K}{K'}(M/N_{CQI})^{K'}(1 - M/N_{CQI})^{K-K'}$. These CQI values are then exploited by the frequency domain PF scheduler, offering a gain of $G(K')$ as compared to the single user case. $G(K') = A \ln K' + B$ is a logarithmic function of the number of users for frequency domain packet scheduling, as presented in Section B.5. Overall, the average frequency domain packet scheduling gain by having best-M report can be approximated by:

$$G_{bestM}(K) = \sum_{K'=1}^{K} P_{K'} G(K') \tag{D.1}$$

Similarly, the reference performance of full CQI report offers a scheduling gain of $G(K)$. The value of M for average best-M report can be choosing according to the following criteria:

- If only one user is available for frequency domain packet scheduling, M should have the value of 0.
- As the number of users available for frequency domain packet scheduling increases, the smallest M that achieves $X\%$ the performance of full CQI report is selected:

$$G_{bestM}(K)/G(K) \geq X \tag{D.2}$$

- If all possible values of M cannot satisfy the requirement of (D.2), $M = \lceil N_{CQI}/2 \rceil$ will be used.

Note that (D.1) is a bit pessimistic in estimating the performance of average best-M. The reason is that, if one CQI group is not considered as "best" by any of the users, it can still be used for transmission to convey some data. However, this is neglected in (D.1). Therefore, X should be set lower than the actual expectation of the best-M performance. Besides the performance gain of adapting M, one should keep in mind that changing the value of M for all users in a cell leads to additional signaling cost. This is even more critical for the finite buffer and bursty traffic cases, where the cell load changes from time to time. Therefore, the trade-off between the performance gain and signaling cost should be considered when deciding how often to change M.

D.2.3 Performance Verification

As presented in Section B.5, in a macro-cell LTE system, $A = 0.11$, $B = 1.10$ and the frequency domain packet scheduling gain saturates at maximum 13 users. With $X = 80\%$ in (D.2), the value of M at different cell load (number of users) for a 10 MHz system is plotted in Fig. D.7.

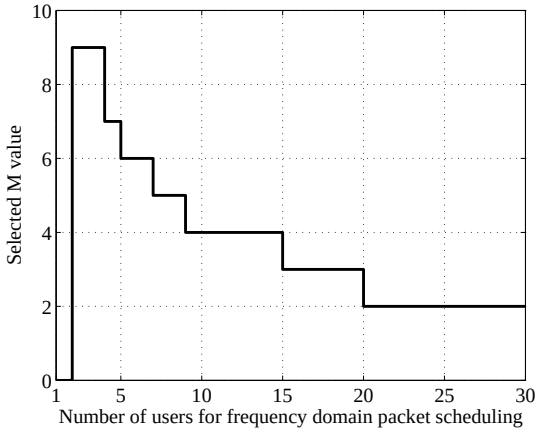


Fig. D.7: Selected M values at different load conditions (number of users, K).

In Fig. D.8, the normalized performance using the load adaptive average best-M with respect to full CQI report is presented. The traditional fixed average best-M is used for reference, and performance is shown with $M = 5$, 9 or 12. It can be seen that a high value of $M = 9$ is beneficial at low load, whereas a low M is preferable when the load is high. Moreover, in the single user case, $M = 0$ is enough to achieve the maximum performance with the lowest overhead. Having M much larger than $N_{CQI}/2$, e.g, 12, is not suggested because it always gives poor performance. Compared to full CQI report, the proposed load adaptive scheme achieves more than 95% average cell throughput and 90% cell edge throughput in all load conditions.

From Fig.D.7 and Fig. D.8, it is observed that using a small value of M is beneficial when a large number of users (higher than 15) are allowed for frequency domain packet scheduling. In practice, due to the control channel limitation and the operation complexity, the number of users that are passed to the frequency domain scheduler usually does not exceed 10. If this is the case, although the optimal value of M varies from 9 to 4, it is observed that a fixed value of $M = 5$ achieves almost the same performance as adaptive best-M. Still, $M = 0$ is preferable if there is only one user per cell.

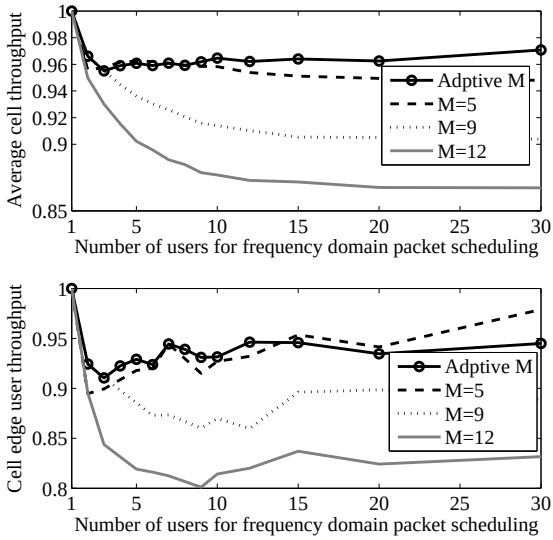


Fig. D.8: Normalized performance of adaptive and fixed best-M with respect to full CQI report.

D.2.4 Summary on Adaptive Average Best-M

It has been shown that the performance of average best-M is sensitive to the cell load. A load adaptive method is proposed to properly select the value of M. When there is only one user per cell, $M = 0$ gives the maximum performance with 94% overhead reduction as compared to full CQI report. When the cell is heavily loaded, a small value of M can be used, e.g. 1 or 2. For medium cell load, M lies between 1 and $\lceil N_{CQI}/2 \rceil$, which can be found using the constraint of (D.2). Performance wise, the adaptive average best-M achieves more than 95% average cell throughput and 90% cell edge user throughput of full CQI report, with at least 69% overhead reduction. While a fixed M may be suitable for a certain range of cell loads, the proposed adaptive best-M is beneficial for all load conditions. The cost is additional overhead for signaling the value of M to the users.

In a realistic LTE system where the number of users available for frequency domain is usually no larger than 10, a fixed value of $M = 5$ achieves very close performance to adaptive best-M. This justifies the use of fixed best-M in [27], which has much reduced signaling overhead and operation complexity than adaptive best-M. Nevertheless, the possibility of using wideband CQI report ($M = 0$) at low load should be considered.

D.3 Hard and Fractional Frequency Reuse

In Chapter 3 the performance of frequency reuse in local area has been investigated, and it is shown that by assigning the same frequency band in different cells, which are far from each other, high inter-cell interference can be avoided while allowing a high utilization of the spectrum. Therefore, frequency reuse is advantageous in such scenarios. In order to distinguish between the Fractional Frequency Reuse (FFR) [152] that will also be discussed in this section, the term Hard Frequency Reuse (HFR) is used to describe the frequency reuse scheme that was used in Chapter 3. FFR is based on the idea of dividing the cell into two zones, and applying a low reuse factor of one in the inner zone with a high reuse factor in the outer zone. This is commonly known as reuse partitioning [44].

In the wide area networks, many studies exist on frequency reuse [44, 45, 82, 153]. In general, a high reuse factor improves the cell edge user throughput at the cost of degraded average cell throughput. The reason for the cell throughput degradation is the unnecessary interference reduction for cell-center users, which at the same time reduces the available spectrum within each cell. To benefit from the improved cell edge user throughput of HFR while trying to minimize its drawback (low average cell throughput), the concept of FFR was developed [152]. It has been adopted in the mobile WiMAX and IEEE 802.20 systems [154, 155], and is also studied for LTE [156]. FFR is usually implemented in OFDMA systems because of the frequency domain flexibility in resource allocation. Indeed, the existing studies for FFR are all based on OFDMA [45, 154–162]. In this section, an alternative approach to realize FFR or HFR is presented, and the performance is evaluated in the downlink transmission of a multi-CC LTE-Advanced system. This approach is based on cell level CC assignment, which does not require the frequency domain scheduling flexibility, and could therefore be used for non-OFDMA systems, e.g. multi-CC CDMA.

D.3.1 Hard Frequency Reuse

HFR first divides the spectrum into several frequency segments, and then applies the same reuse factor for all. In a multi-CC system, each segment corresponds to one CC, and different frequency reuse factors can be obtained by changing the number of CCs assigned to each cell.

Let N, M denote the total number of aggregated CCs and the number of CCs assigned to each cell. The indices for the assigned CCs to cell i can be calculated as:

$$C(i) = mod\{[i, i+1, \ldots, i+M-1], N\} \tag{D.3}$$

And the corresponding reuse factor is

$$r_i = \sum_{n=0}^{N-1} W_n / \sum_{m \in C(i)} W_m \tag{D.4}$$

where W_n is the bandwidth for CC n. When the aggregated CCs have equal bandwidth, $r = N/M$ for all cells. Fig. D.9 gives an example of achieving frequency reuse-3 by assigning each cell with only one out of the three aggregated CCs.

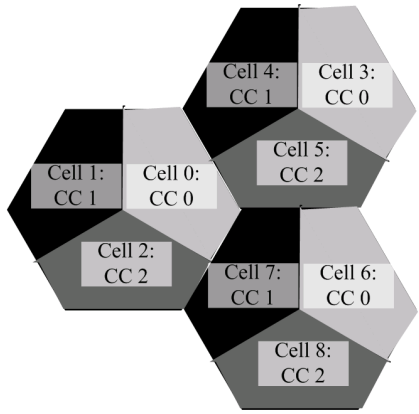


Fig. D.9: HFR in a multi-CC system. Each cell is assigned with one out of three CCs, resulting in reuse-3 mode.

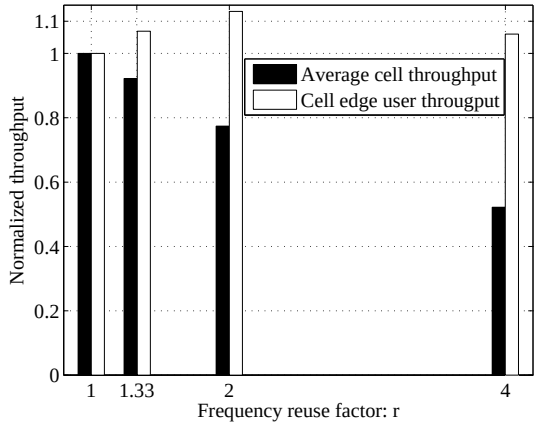


Fig. D.10: Performance of HFR with different reuse factors.

For the performance evaluation, the system setup is as listed in Table A.3. It assumes four aggregated CCs, with 10 MHz bandwidth each. The full buffer

traffic model is used, with 10 users per cell. Different reuse factors are obtained by changing the number of CCs assigned to each cell, according to (D.4). As expected, Fig. D.10 shows that HFR can improve the coverage performance for reuse factor up to 2. A further increase in reuse factor gives a loss in both cell throughput and coverage. The reason is that with HFR, the available bandwidth for each cell is inversely proportional to the reuse factor. This penalizes the system performance severer than the gain in interference reduction. Note that the performance is normalized with respect to the case when frequency reuse one is applied to all CCs.

D.3.2 Fractional Frequency Reuse

Similar to HFR, FFR can also be realized by the CC assignment in the cell level. It works by dividing the CCs into two groups, one group is universally used by all cells to serve users in the inner zone; the other group has a reuse factor higher than one, it will serve the outer zone users who suffer from high inter-cell interference. The division of the inner and outer zone users is based on the channel quality measurement, as in [158]. An example of four CCs, where one CC is used in all cells and the rest of the CCs follow a reuse-3 manner is depicted in Fig. D.11.

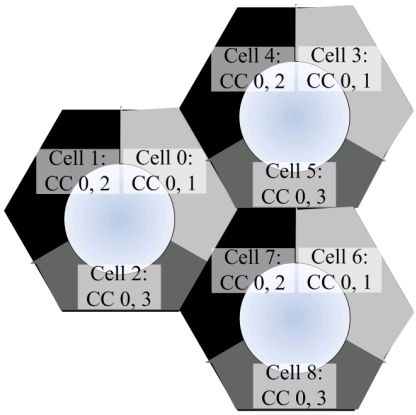


Fig. D.11: FFR in a multi-CC system. Each cell is assigned with two out of four CCs. One follows university reuse, the others follow reuse factor of 3.

FFR improves the SINR for users in the outer zone. The users close to the base station maintains their SINR and achieves high spectrum utilization. The performance of FFR is sensitive to the division of inner and outer zone users. A portion of the users with poor channel quality will be considered as outer zone users, and the rest are inner zone users. This is controlled by a tunable parameter τ, which is

the ratio of outer zone users over the total users. In the simulation different values of τ are used, with the aim to find the optimal division of inner and outer zone users. In order to compare with HFR, the same aggregation pattern of 4 equal size CCs is used. This allows the following two FFR configurations:

FFR-config.1: One CC with reuse factor of 1, plus three CCs with reuse 3.

FFR-config.2: Two CCs with reuse factor of 1, plus two CCs with reuse 2.

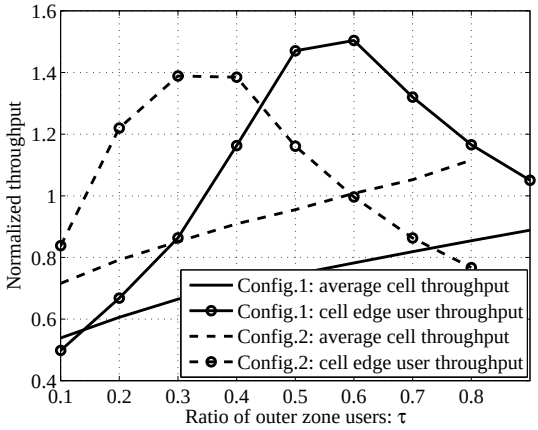


Fig. D.12: Performance of FFR with different reuse patterns and different ratios of outer zone users, normalized with respect to reuse-1.

The results of FFR are summarized in Fig. D.12 for both normalized average cell throughput and cell edge user thoughput. From this figure it can be seen that the highest cell edge user throughput for FFR-config.1 is obtained when 60% users are considered to be in the outer zone, which is 51% higher than reuse 1. Meanwhile, the loss in average cell throughput is 21%; for FFR-config.2, the cell edge user throughput can be improved by as much as 39%, with only 9% degradation in cell throughput. It can also be seen that FFR-config.2 offers a better trade-off between cell throughput and coverage as compared to FFR-config.1, in the sense that it achieves similar gain in cell edge user throughput with better average cell throughput.

Comparing Fig. D.10 with Fig. D.12, it can be seen that with the proper division of inner and outer zone users, FFR offers better performance than HFR in both average cell and cell edge user throughput. Therefore it is a better solution than HFR in wide area networks.

D.3.3 Summary on Frequency Reuse in Wide Area Networks

Based on the simulation results, HFR achieves very little gain in cell edge user throughput but severely penalizes the average cell throughput than reuse-1. Therefore it is not recommended for wide area systems. This conclusion is in agreement with the wide area studies in the open literature. However, FFR has several attractive potentials and worth further investigation: first, it offers the trade-off between average cell throughput and cell edge user throughput; secondly, it reduces the signaling overhead and power consumption by allowing the cells/users to operate with a narrower bandwidth than reuse-1. The drawback of FFR is that it requires a proper frequency plan, which may be difficult to perform in real networks. Also, a proper division of the users on different frequency bands is required.

It is worth mentioning that similar trade-off between average cell (or user) throughput and cell edge user throughput can also be obtained with GPF. However, GPF fails to improve the cell edge user throughput with the finite buffer model, because the good and poor users are competing for resources with each other. FFR avoids the competition between these two kinds of users by serving them with different CCs, therefore can improve the cell-edge user throughput even for the finite buffer case.

D.4 User Velocity Dependent CC Assignment

D.4.1 Introduction

The performance of a wireless communication system is sensitive to the user velocity as well as the carrier frequency. Let v denote the velocity, f_c for the carrier frequency, $c = 3.0 \times 10^8$ m/s is the speed of light, the Doppler frequency is $f_m = vf_c/c$ [12]. It is also known that the channel coherence time, which is a statistical measure of the time duration over which the channel impulse response is invariant, is inversely proportional to the Doppler frequency [12]. When the channel feedback delay is the same on all CCs, a high velocity or high carrier frequency therefore leads to less reliable feedback, and vice versa. Also, the path loss is sensitive to the carrier frequency, and hence CCs could have different coverage areas [9].

In the previous studies, the contiguous carrier aggregation has been considered. However, LTE-Advanced should also support carrier aggregation over non-contiguous CCs, e.g. one in 900 MHz and the other in 3.5 GHz [9, 20]. Also, the users inside

a macro-cell may have different velocities. If the high velocity users are served by the CC with high frequency, the coherence time will be very short. This reduces the reliability of the channel feedback and prevents the system from fully exploiting the frequency domain diversity. Therefore, it is suggested to perform the CC assignment with the awareness of the user velocity, putting the high velocity users on the CC with low frequency and low velocity users to the CC with high frequency.

D.4.2 Performance of Velocity Dependent CC Assignment

The performance is evaluated with the carrier aggregation pattern of 2 CCs, one at 900 MHz and the other at 3.5 GHz. The bandwidth is 10 MHz for each CC. The macro-cell scenario as described in Table A.3 is used here, with periodical feedback every 5 ms and a feedback delay of 6 ms. There are 10 full buffer users in each cell. 5 of them have a low velocity of 3 kmph, and the rest have a higher velocity of 10 kmph. The performance for the speed dependent CC assignment is evaluated and compared to the velocity-blind assignment, with one CC assigned to each user.

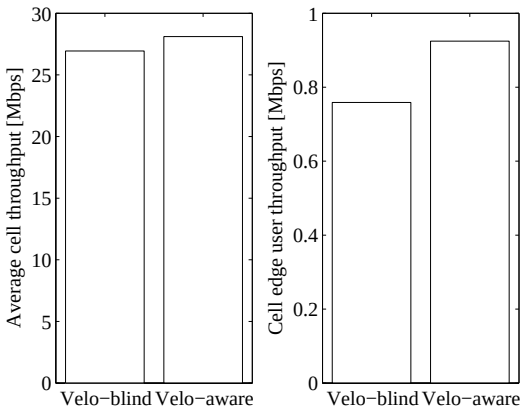


Fig. D.13: Performance with non-contiguous CCs and different user velocity.

Fig. D.13 shows the performance with the two assigning methods. The LL balancing is used, resulting in 5 users per CC. With the given system setup, velocity-dependent assignment achieves 4.5% gain in average cell throughput and 21.7% gain in cell edge user throughput, as compared to velocity-blind assignment.

D.4.3 Summary on the Velocity Dependent CC Assignment

If the aggregated CCs have different center frequency, assigning high velocity users to the lower frequency CC may be beneficial to improve the system performance. This section aims to show the possible gain by having the velocity dependent CC assignment. For this purpose, the simplest case with the same number of low velocity and high velocity users is assumed, and these users are evenly divided on 2 CCs. The obtained result indicates that other than the most important factors of CC load and the QoS requirement, the user velocity could also be considered in the CC assignment.

D.5 Conclusion

Several techniques to improve the system performance of an LTE(-Advanced) system have been investigated in this appendix. Some of them can be applied to both single and multi-CC systems, e.g. the GPF scheduling and load adaptive average best-M; others require the existence of multiple CCs, e.g., the cell-level CC assignment to achieve hard/fractional frequency reuse and the user velocity dependent CC assignment.

Based on the simulation results, it can be concluded that as compared to traditional PF, the GPF scheduler offers more flexibility in the resource allocation and is more preferable for the resource sharing among best effort users. Moreover, the behavior of GPF is dependent on the traffic models, e.g., it can trade off between average cell (user) throughput and cell edge user throughput when the full buffer traffic model is used; with the finite buffer model, a small value of β can improve the average user throughput without losing in cell edge user throughput. The average best-M can be used to reduce the CQI feedback in each CC. Although the optimal value of M is sensitive to the cell load, a fixed value of $M = 5$ achieves nearly the best performance with realistic LTE assumptions. Therefore it is suggested to use a fixed value of M in LTE, with the exception that $M = 0$ could be used in the single user case. HFR is found to be inefficient in the wide area scenario, and FFR could potentially be used to improve the cell edge user throughput with very small loss in average cell throughput. Finally, the velocity dependent CC assignment is investigated, and it is suggested that other than the CC load and user QoS requirement, the user velocity could also be considered (with low weighting) when performing the CC assignment in the non-contiguous carrier aggregation scenario.

Appendix E

Supplementary Results for Previous Chapters

This chapter provides the supplementary results and analyses for the previous chapters. In Section E.1, the impact of network time synchronization on a D-TDD system is shown; Section E.2 provides the extension of the FiDCA algorithm (which was developed in Chapter 3) to support high frequency reuse factors; Section E.3 summarizes the performance of CQI reduction with the full buffer traffic model and complements the analysis in Section 5.2.

E.1 Dynamic TDD with Different Levels of Network Time Synchronization

The impact of network time synchronization on a S-TDD system has been evaluated in Chapter 2, which assumes the same uplink to downlink ratio of 50-50% in all cells. To accommodate for varying traffic patterns across different cells, different subframe partitioning between uplink and downlink transmissions may be used, leading to the D-TDD mode. While providing better adaptation to the traffic requirement per cell, D-TDD is subject to uplink-downlink mutual interference even with full synchronization. As a consequence, the impact of network time synchronization on a D-TDD system may be different from a S-TDD system.

The objective of this section is to evaluate the performance of D-TDD with different synchronization levels. The evaluation is carried out based on simulations, following the guideline in Section A.1. During the simulation of the D-TDD mode, the ratio between uplink and downlink transmissions is taken as randomized. Let Υ denote the resource sharing of the downlink transmission over the total available resources. It is assumed that Υ uniformly distributes between 0 and 1, and hence the average uplink and downlink traffic volume is the same. Fig. E.1 gives an example of the transmission pattern in a local area network with 4 base stations.

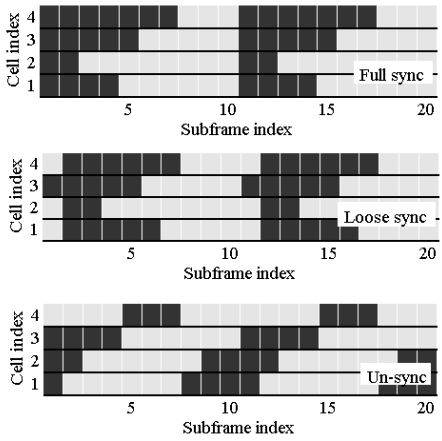


Fig. E.1: Transmission patterns with D-TDD and different synchronization levels.

E.1.1 Performance of Dynamic TDD in Small Indoor Scenarios

The performance presented in this subsection is evaluated in the small indoor home and office scenarios, with 4 cells each. Similar results have also been obtained in the large scenarios, which maintain similar trends as in the small scenarios and are therefore not shown.

Fig. E.2 summarizes the CDF of the downlink user throughput of a D-TDD system with different synchronization levels, where the left subplot is for the home scenario and the right subplot is for the office scenario. As shown, the performance is hardly affected by the synchronization level, because even with full synchronization, D-TDD cannot offer full isolation between downlink and uplink, as indicated in the top subplot of Fig. E.1.

From the uplink performance shown in Fig. E.3, it is observed that achieving

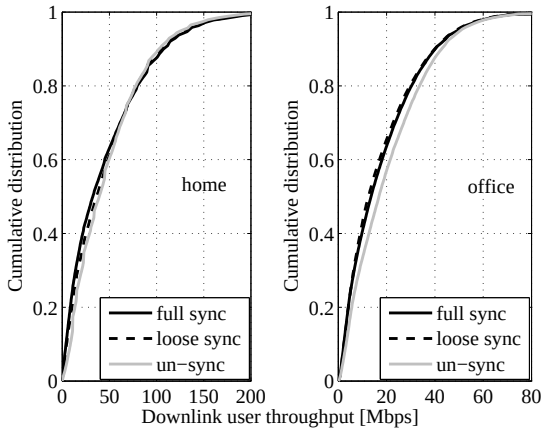


Fig. E.2: CDF of downlink user throughput with different synchronization levels in a D-TDD system.

network synchronization is beneficial. With the considered system setup, full synchronization offers 81% higher average cell throughput in the home scenario, and 93% higher in the office scenario than without synchronization. Meanwhile, it also reduces the percentage of users that cannot be served from around 50% to 10% in both home and office scenarios.

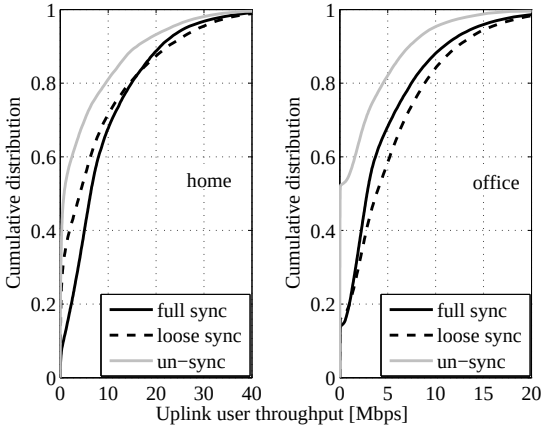


Fig. E.3: CDF of uplink user throughput with different synchronization levels in a D-TDD system.

E.1.2 Conclusion

Based on the results shown in this section, achieving network time synchronization significantly improves the uplink performance of a D-TDD system compared to the case without synchronization. This is similar to the S-TDD system and justifies the need for synchronizing the transmission time across the local area base stations. However, D-TDD is less sensitive to the level of network synchronization than S-TDD. It has similar downlink performance for the different synchronization levels, and the difference in the uplink performance is smaller than S-TDD.

The full buffer traffic model has been used for the performance evaluation of both S-TDD and D-TDD. With this traffic model, the benefit of D-TDD to adapt to the varying uplink/downlink traffic load cannot be seen. A proper traffic model is required for a fair comparison between the two TDD modes, which is left for future work.

E.2 Extension of FiDCA to Support Any Frequency Reuse Factor

In the studied local area scenario, a reuse factor of 2 offers the best performance among all possible reuse factors. To avoid the necessity of network planning, Chapter 2 develops an algorithm called Firefly inspired Dynamic Channel Allocation (FiDCA), which operates in a self-organized manner and aims to achieve the frequency reuse-2 pattern. It offers significant gain than random channel allocation or universe frequency reuse (reuse-1), and fulfills the original purpose of approaching the performance of reuse-2. Therefore, it has great potential of being used in real systems for the dynamic channel allocation.

Although reuse-2 offers the best performance in the studied local area scenarios, there might be cases when a higher reuse factor is beneficial, e.g., if the base stations are deployed in a very dense manner with even smaller coverage and hence higher inter-cell interference than the cases considered in this thesis. Another case that desires a higher reuse factor is, when an operator has multiple (denoted by N) aggregated CCs and legacy base stations, each legacy base station supports only one CC, leading to a reuse-N manner. Considering these reasons, it is worth the effort to extend the FiDCA algorithm to support any reuse factors. The extension is described in the following.

E.2.1 Description of the Extended FiDCA Algorithm

The extended FiDCA still operates based on the status indicator s_i, as the original one described in Section 3.3.1. However, it extends s_i to a vector $\mathbb{S}_i = \{s_{i,1}, \ldots, s_{i,N}\}$, where $0 < s_{i,n} < 1$ is the status indicator for cell i on the n^{th} channel. $s_{i,n}$ increases if channel n has better quality (average SINR) than the current channel, or decreases towards 0 if the current channel has better channel quality. When $s_{i,n}$ reaches the level of 1, the channel allocation will be updated by switching to channel n. Afterwards the status indicator will be reset to a common initial point for all candidate channels. The flowchart of the extended FiDCA is presented in Fig. E.4, and the operations for each step are provided below.

Step 1: Initialization. Randomly selects a channel $H_i \in \{1, 2, \ldots, N\}$ for cell i among all possible channels. The un-selected channels are the candidate channels that the cell may switch to, they are denoted by the set of $\mathbb{D}_i = \{1, \ldots, H_i - 1, H_i + 1, \ldots, N\}$. Assign a random initial status $\mathbb{S}_i$ to each candidate channel, which uniformly distributes between 0 and 1, and i is the cell index. All the members in the set $\mathbb{S}_i$ has the same value, i.e, $s_{i,m} = s_{i,n}\ \forall\ m, n \in \mathbb{D}_i$. The value of $s_{i,n}$ indicates the satisfaction of cell i about its current channel allocation H_i as compared to channel n. If this value is close to 1, it means the cell prefers channel n more than the current channel. If this value is close to 0, it means the cell is satisfied with the current allocation and is reluctant to switch.

Step 2: At each update slot, one cell will compare the average SINR level between the current (γ_c) and all candidate channels ($\gamma_n, n \in \mathbb{D}_i$). Based on the difference (in dB), it will adjust $\mathbb{S}_i$ according to the following rules:

a) $\gamma_c > \gamma_n$: favors current channel, jump towards the value 0.

b) $\gamma_c \leq \gamma_n$: favors channel n, jump towards the value 1.

For each $n \in \mathbb{D}_i$, the value of $s_{i,n}$ after the jump is:

$$s'_{i,n} = (1 - \alpha)s_{i,n} + \alpha(s_{i,n} + s)/2 \tag{E.1}$$

In (E.1), $s = 0$ if the jumping favors the current channel, otherwise $s = 2$; $(s_{i,n} + s)/2$ is the mid-point between the current status and the desired operating point, which is similar to the one in [92]; α controls the speed of the jumping, it is based on the SINR gap between the two channels, a pre-defined maximum jumping step β, and the maximum SINR gap G:

$$\alpha = \beta min\{|\gamma_c - \gamma_n|/G, 1\} \tag{E.2}$$

Step 3: Find the maximum value of $\mathbb{S}_i$ (denoted by s_i^{max}) and the corresponding channel index n^{max} that has the maximum value. If $s_i^{max} > 1$:

1. switch to channel n^{max}.
2. Update the candidate channel set $\mathbb{D}_i$ and reset the status indicator values in $\mathbb{S}_i$ to be $2 - s_i^{max}$ for all candidate channels.

Otherwise, wait until the next update slot.

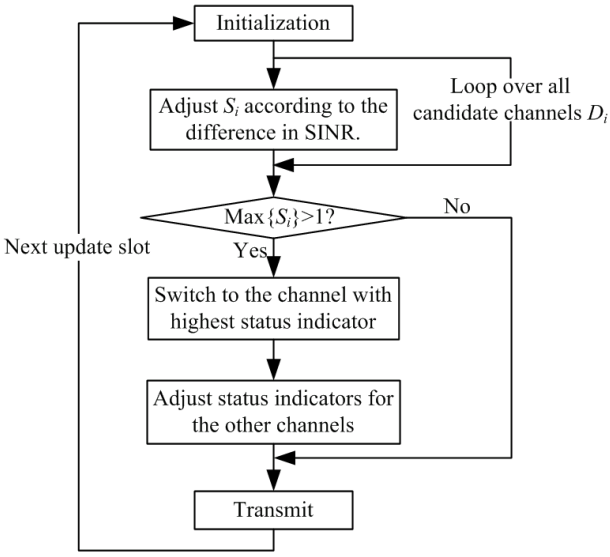


Fig. E.4: Flowchart of the extended FiDCA.

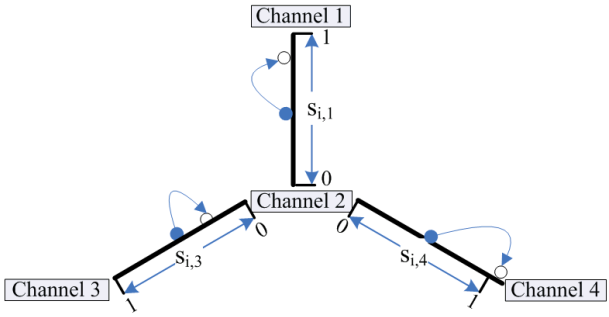


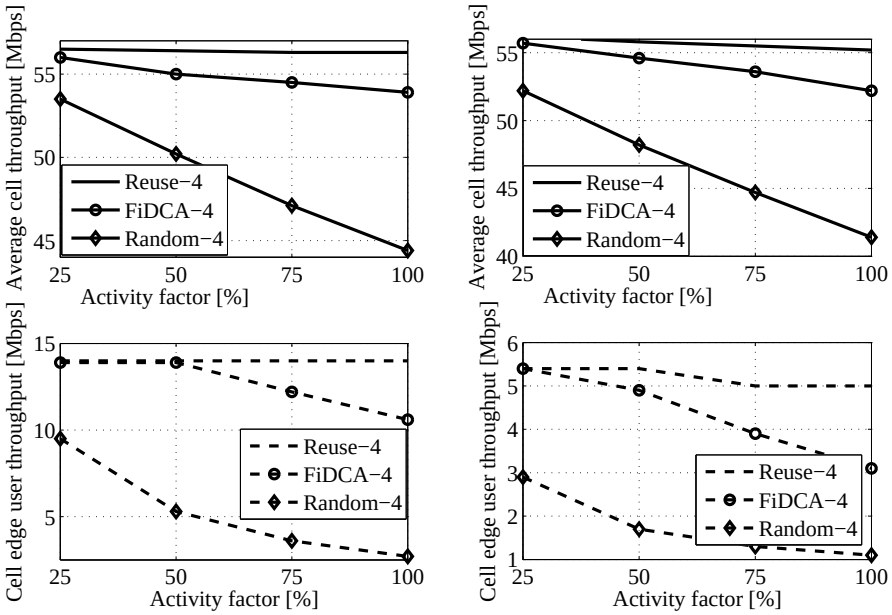
Fig. E.5: An example of FiDCA to achieve reuse-4.

An example of FiDCA with 4 channels (reuse-4) is given in Fig. E.5. Channel 2 is currently allocated to cell i. It has lower SINR than channel 1 and 4, but higher SINR than channel 3. As a consequence, $s_{i,1}$ and $s_{i,4}$ increase as time goes by and $s_{i,3}$ decreases. In this example, channel 4 has the highest SINR and hence $s_{i,4}$ jumps with larger step than the others. With the same initial values (the same 'distance' to the candidate channels), channel 4 will be selected after a

certain time, and a new round of evolution will start. The system stabilizes when all cells are satisfied with their channel allocation, and $s_{i,n} = 0^+$ for all candidate channels.

E.2.2 Performance of Extended FiDCA

To evaluate the performance of extended FiDCA, the large indoor home and office scenarios as described in Section A.1 are used. The evaluation is done for the case of reuse 4, with both planned and random reuse-4 being the reference. Different cell activity factors have been considered, which correspond to the probability of having an active base station within each cell.

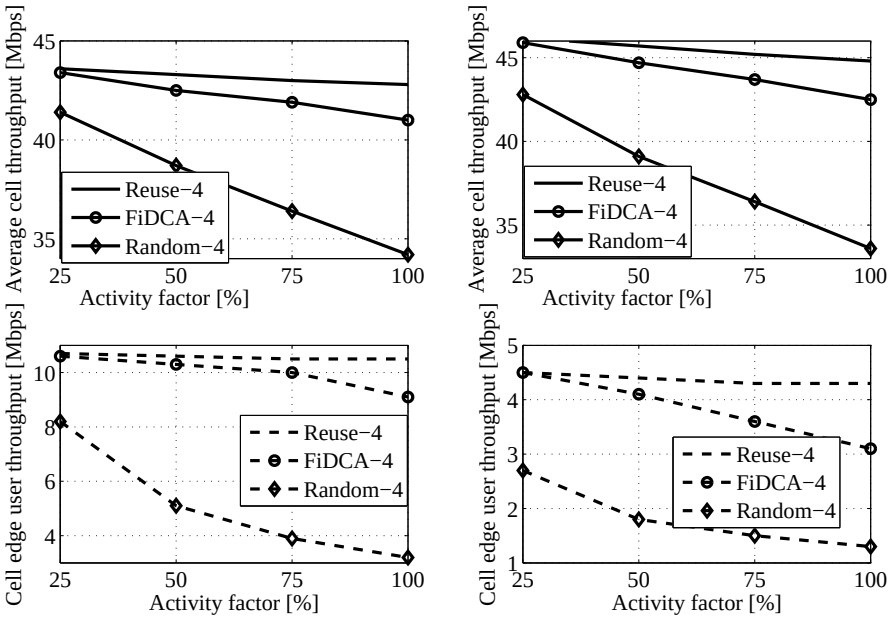


(a) Home scenario

(b) Office scenario

Fig. E.6: Downlink performance of FiDCA towards Reuse-4.

Fig. E.6 and Fig. E.7 show the average cell throughput and cell edge user throughput of FiDCA and planned/random reuse-4 in the downlink and uplink of the indoor scenarios, respectively. In each of these two figures, the left subplot is for the performance in the home scenario, and the right subplot is for the performance in the office scenario. As can be seen from these figures, the extended FiDCA



(a) Home scenario

(b) Office scenario

Fig. E.7: Uplink performance of FiDCA towards Reuse-4.

achieves similar performance as planned frequency reuse, which is much better than random channel allocation. The relative performance of FiDCA as compared to frequency reuse decreases with the activity factor, indicating sub-optimal frequency planning when a large number of cells exist.

E.2.3 Conclusion

This section extends the previously developed FiDCA algorithm to support any frequency reuse factor, unlimited to reuse-2. It outperforms random frequency reuse and approaches the performance of planned frequency reuse. Meanwhile, it maintains the operational simplicity of the original FiDCA and is therefore a good candidate for the frequency planning in the uncoordinated local area network. Although orthogonal channels have been used in the simulation, it also supports overlapping channels and leads to fractional frequency reuse.

E.3 CQI Reduction With Full Buffer Traffic Model

In Chapter 5, the performance of Channel Quality Information (CQI) reduction techniques has been analyzed by using the finite buffer model. In this chapter, the simulation results generated with the full buffer model are provided. The purpose is to verify the robustness of the CQI reduction techniques with respect to different traffic models. Therefore, the simulation assumptions are the same as in Chapter 5, except for the traffic model. Different load conditions are obtained by varying the number of users per cell from 1 to 10 in the simulation.

E.3.1 Performance With Intra-CC CQI Reduction

The performance with different intra-Component Carrier (CC) CQI reduction techniques are presented in Fig. E.8, where the upper plot is for the average cell throughput, and the lower plot is for the cell edge user throughput. The performance is obtained in a 4×10 MHz system, with each user scheduled on all CCs.

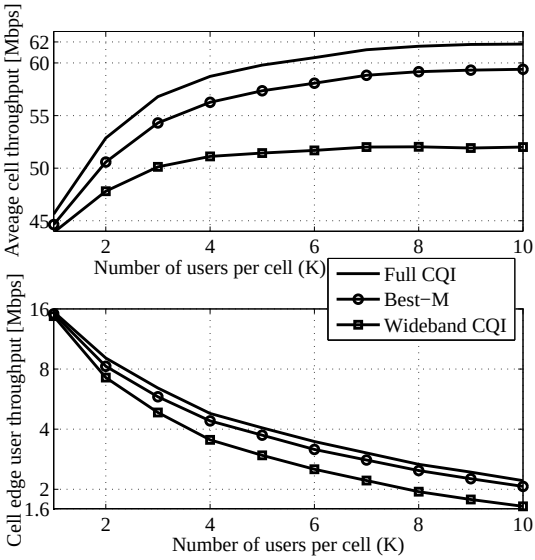


Fig. E.8: Performance with intra-CC CQI reduction.

As shown in Fig. E.8, different feedback schemes have the same performance when there is a single user per cell. This is due to the fact that the user is anyway

scheduled on the whole bandwidth, and link adaptation is performed based on the wideband CQI. As the number of users increases, the system will benefit from multi-user diversity. Even with wideband CQI report, having multiple users makes it possible to schedule a user when its average channel quality is good, and hence the cell throughput slightly increases with the number of users in a cell. In addition to multi-user gain, full CQI and best-M also provide the scheduler with frequency domain diversity, which further improves the throughput as compared to wideband CQI. Their throughput versus the number of users follows a logarithmic shape, which is also witnessed in Section B.5. Using best-M or full CQI report, the performance with 10 users is 30∼33% higher than the single user case, which is close to the estimation according to (B.21): $0.11 \ln 10 + 1.1 = 1.35$. In cell edge, the multi-user gain of having 10 users is 1.38 in simulation, while the estimation using (B.22) is $0.15 \ln(10) + 1.05 = 1.40$.

With respect to full CQI report, best-M causes around 5% loss in average cell throughput and less than 10% loss in cell edge user throughput. As to wideband CQI, its relative performance decreases with the number of users, due to the lack of frequency domain diversity. When 10 users exist, the loss is 16% in average cell throughput, and 27% in cell edge user throughput.

E.3.2 Performance With Inter-CC CQI Reduction

It has been found in the previous study that best-M achieves similar performance as full CQI report, with significant overhead reduction. Therefore, using best-M in all CCs is taken as the reference case for the study of inter-CC CQI reduction. Different inter-CC CQI reduction techniques are evaluated using the full buffer model, and the results are shown in this section.

E.3.2.1 Performance With B&W

The performance of Best and Wideband CQI (B&W) with different number of CCs (N_{BM}) using best-M report is summarized in Fig. E.9. Like in the intra-CC CQI reduction, it is observed that reporting only the wideband CQI is enough if there is only one user per cell. When the number of users increases, the wideband CQI report prevents the packet scheduler from exploiting the frequency domain diversity, and the performance is worse than using best-M in all CCs. As an example, when 10 users exist in each cell, B&W with $N_{BM} = 1$ has 4.6% lower average cell throughput and 6.8% lower cell edge user throughput than best-M in all CCs. The performance of B&W can be slightly improved by having more CCs with best-M report (increase N_{BM}), at the cost of high additional feedback overhead (see Fig. 5.2). It is therefore suggested to use B&W only at low cell load,

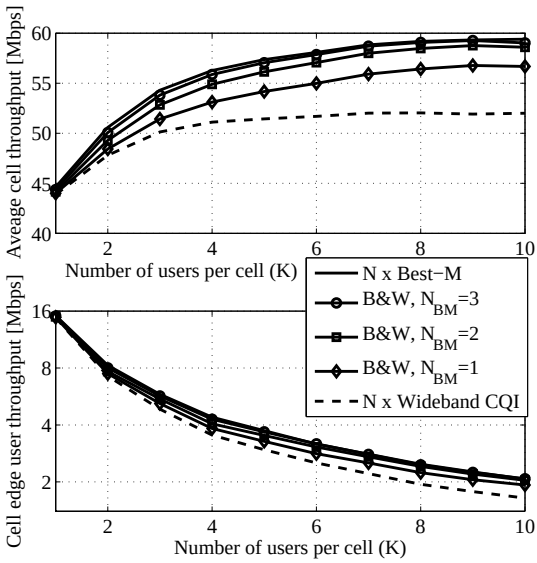


Fig. E.9: Performance with inter-CC CQI reduction: B&W.

with $N_{BM} = 0$ ($N\times$Wideband CQI) or 1.

E.3.2.2 Performance With BCC

Fig. E.10 shows the performance of Best-CC (BCC) with different number of users per cell. As can be seen from this figure, when there is a single user per cell, BCC with $N_{BM} = 1$ achieves slightly higher than 1/4 the performance of using best-M in all CCs. This is due to the fact that only one CC is used and the throughput suffers from inefficient resource utilization. BCC is efficient at medium to high load, and the trade-off between downlink performance and uplink overhead can be made by tuning the value of N_{BM}.

When the number of users is high enough, BCC offers enough frequency domain diversity gain with only one CC per user. It achieves similar performance as best-M in all CCs, with very low feedback overhead. It should be noted that in the simulation, the worst case is considered, where the surrounding cells are transmitting all the time on all CCs. At low load, this causes higher interference than it should be, and leads to poorer performance. This problem does not exist for the high load case when all CCs are utilized.

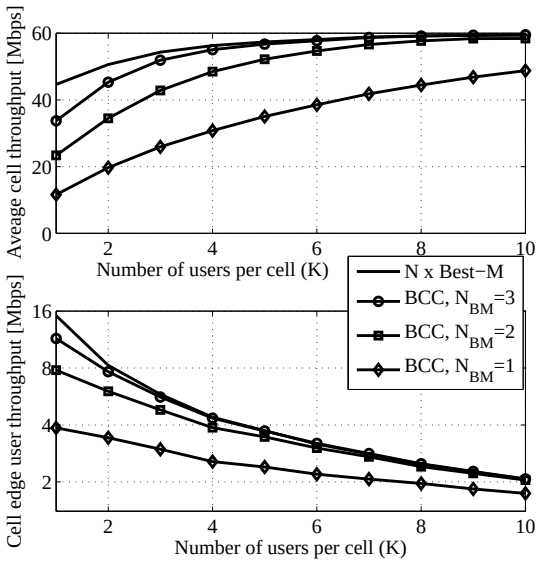


Fig. E.10: Performance with inter-CC CQI reduction: BCC.

E.3.2.3 Performance With BCL

The simulation of BCL has the same problem of over-estimating the interference at low load, i.e., when $KN_{BM} < N$. However, its performance with realistic interference condition is similar to the frequency reuse case. According to Fig. D.10, the average cell throughput decreases by maximum 50% when only one CC is occupied, and the cell edge user throughput slightly benefits (maximum 13%) from assigning less number of CCs to each user. In general, the performance of BCL is similar to BCC, as can be seen from Fig. E.11, and differs from BCC only in the following aspects:

1. The CC assignment is performed centrally at the base station, trying to distribute the users evenly on the available CCs. Therefore, the possibility of having unused CCs is much lower than BCC. It offers better performance than BCC at low load, except for the single-user case.

2. The assigning of CCs to each user is fixed, and hence BCL cannot benefit from the frequency domain diversity across the CCs. Starting from very load, the average cell throughput increases linearly with the number of occupied CCs, until every CC is utilized. Meanwhile, the cell edge user throughput remains constant.

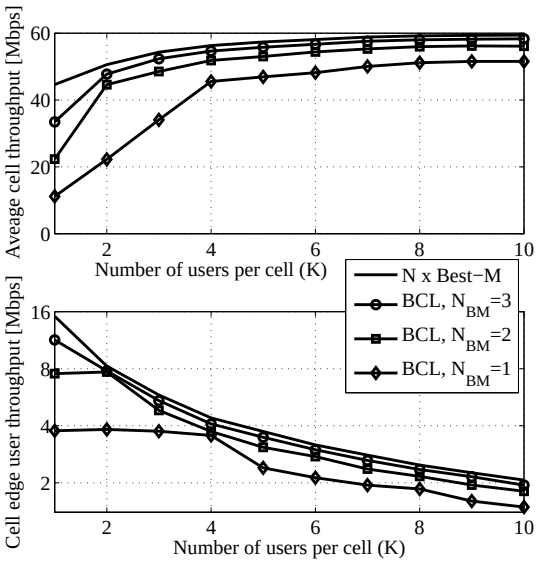


Fig. E.11: Performance with inter-CC CQI reduction: BCL.

E.3.3 Load Adaptive CQI Reduction

The results shown in this chapter also indicate the necessity to select a proper CQI reduction technique based on the load condition, hence a similar approach as presented in Section 5.2.4 can be used. However, In Section 5.2.4, the calculation of the number of CCs for each LTE-Advanced user is applicable only for the finite buffer model. If the full buffer model is used, all users arrive at the same time, and the base station knows the number of LTE-Advanced and Rel'8 users at each simulation run, the following equation should be used to calculate the needed CCs per LTE-Advanced user:

$$N_{BM} = min\left\{max\left\{1, \left\lceil \frac{NT - (K - K_\alpha)}{K_\alpha} \right\rceil \right\}, N \right\} \tag{E.3}$$

In (E.3), N is the number of aggregated CCs; T is the number of users that is able to achieve enough multi-user gain; K is the total number of users; K_α is the number LTE-Advanced users; $K - K_\alpha$ is the number of LTE-Rel'8 users. The goal of (E.3) is the same as (5.5), i.e., to ensure T users per CC whenever possible.

E.3.4 Conclusion

In this chapter the supplementary results for CQI reduction with the full buffer traffic model are presented. In general, the loss in cell edge user throughput when using CQI reduction is slightly lower than that obtained in Chapter 5, where the finite buffer model is used. The reason is that, with the full buffer model, users are kept alive during the simulation time, therefore reducing the CQI only leads to lower frequency domain diversity. For the finite buffer case, other than the loss in frequency domain diversity, reducing CQI also gives rise to the user alive time. As a consequence, more users will exist in a system and compete for resources with each other.

Despite the small difference in the gain/loss values, the general trends are the same for both traffic models. Therefore, it verifies the CQI reduction techniques can be applied for different traffic models.

List of Acronyms

1G First Generation

2G Second Generation

3G Third Generation

3GPP Third Generation Partnership Project

ACCS Autonomous Component Carrier Selection

ACK Acknowledgement

AMPS Advanced Mobile Phone System

AP Access Point

BCC Best-CC

BCL Balanced CC Load

BLER Block Error Rate

CC Component Carrier

CDF Cumulative Distribution Function

CDMA Code Division Multiple Access

CQI Channel Quality Information

CSG Closed Subscriber Group

CSI Channel State Information

DCA Dynamic Channel Allocation

DS^3 Dynamic Spectrum Sharing with Selfishness

D-TDD Dynamic TDD

EESM Exponential Effective SIR Mapping

EVM Error Vector Magnitude

FCA Fixed Channel Allocation

FDD Frequency Division Duplexing

FFR Fractional Frequency Reuse

FiDCA Firefly inspired Dynamic Channel Allocation

FR Fair Resource

FSU Flexible Spectrum Usage

GPF Generalized Proportional Fair

GPS Global Positioning System

GSM Global System for Mobile communication

HARQ Hybrid Automatic Repeat Request

HFR Hard Frequency Reuse

HSDPA High Speed Downlink Packet Access

IEEE Institute of Electrical and Electronics Engineers

IRC Interference Rejection Combining

ILLA Inner-Loop Link Adaptation

IMT International Mobile Telecommunications

ITU-R International Telecommunications Union - Radio Communication Sector

LL Least Load

LOS Line of Sight

LTE Long Term Evolution

MCS Modulation and Coding Scheme

MH Mobile Hashing

MIMO Multiple Input Multiple Output

MRC Maximal Ratio Combining

NACK Non-acknowledgement

NLOS Non-Line of Sight

OFDM Orthogonal Frequency Division Multiplexing

OFDMA Orthogonal Frequency Division Multiple Access

OLLA Outer-Loop Link Adaptation

OSG Open Subscriber Group

PAPR Peak to Average Power Ratio

PDF Probability Distribution Function

PF Proportional Fair

PMI Precoding Matrix Indicator

PRB Physical Resource Block

PS Packet Scheduling

QAM Quadrature Amplitude Modulation

QoS Quality of Service

QPSK Quadrature Phase Shift Keying

RI Rank Indication

RR Round Robin

RRM Radio Resource Management

S-TDD Static TDD

SC-FDM Single-carrier Frequency Division Multiplexing

SIMO Single Input Multiple Output

SINR Signal to Interference plus Noise Ratio

SISO Single Input Single Output

STBC Space-Time Block Code

TDD Time Division Duplexing

TTI Transmission Time Interval

WCDMA Wideband Code Division Multiple Access

WiMAX Worldwide Interoperability for Microwave Access

WLAN Wireless Local Area Network

ZF Zero Forcing

Bibliography

[1] W. R. Young, "Advanced mobile phone service: Introduction, background, and objectives," *Bell Systems Technical Journal*, vol. 58, pp. 1–14, Jan 1979.

[2] J. Padgett, C. Günther, and T. Hattori, "Overview of wireless personal communications," *Communications Magazine, IEEE*, vol. 33, pp. 28–41, Jan 1995.

[3] A. Maloberti, "Radio transmission interface of the digital panEuropean mobile system," in *Vehicular Technology Conference, 1989, IEEE 39th*, vol. 2, pp. 712–717, May 1989.

[4] T. Halonen, J. Romero, and J. Melero, *GSM, GPRS and EDGE performance: evolution towards 3G/UMTS*. John Wiley & Sons, 2002.

[5] H. Honkasalo, K. Pehkonen, M. Niemi, and A. Leino, "WCDMA and WLAN for 3G and beyond," *Wireless Communications Magazine, IEEE*, vol. 9, pp. 14–18, Apr 2002.

[6] H. Holma and A. Toskala, *WCDMA for UMTS - HSPA evolution and LTE, fifth edition*. John Wiley & Sons, 2010.

[7] 3GPP, "UE radio access capabilities (release 1999)," *Tech. Spec. 25.306 V3.10.0*, Dec 2003.

[8] IEEE, "Part 16: Air interface for fixed broadband wireless access systems," *Std. 802.16-2004 (Revision of IEEE Std. 802.16-2001)*, Oct 2004.

[9] 3GPP, "Evolved universal terrestrial radio access (E-UTRA) and evolved universal terrestrial radio access network (E-UTRAN); overall description (release 10)," *Tech. Spec. 36.300 V10.0.0*, Jun 2010.

[10] D. Astély, E. Dahlman, A. Furuskar, Y. Jading, M. Lindstrom, and S. Parkvall, "LTE: the evolution of mobile broadband," *Communications Magazine, IEEE*, vol. 47, pp. 44–51, Apr 2009.

[11] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G evolution: HSPA and LTE for mobile broadband*. Elsevier, 2007.

[12] T. S. Rappaport, *Wireless communications principles and practice, second edition*. Prentice Hall, 2001.

[13] ITU-R, "Framework and overall objectives of the future development of IMT-2000 and systems beyond IMT-2000," *Recommendation M. 1645*, Jun 2003.

[14] A. Hashimoto, H. Yoshino, and H. Atarashi, "Roadmap of IMT-advanced development," *Microwave Magazine, IEEE*, vol. 9, pp. 80–88, Aug 2008.

[15] S. Parkvall, E. Dahlman, A. Furuskär, Y. Jading, M. Olsson, S. Wänstedt, and K. Zangi, "LTE-Advanced - evolving LTE towards IMT-Advanced," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pp. 1–5, Sep 2008.

[16] S. Ahmadi, "An overview of next-generation mobile WiMAX technology," *Communications Magazine, IEEE*, vol. 47, pp. 84–98, Jun 2009.

[17] R. Marks, J. Costa, and B. Kiernan, "The evolution of WirelessMAN," *Microwave Magazine, IEEE*, vol. 9, pp. 72–79, Aug 2008.

[18] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); further advancements for E-UTRA physical layer aspects (release 9)," *Tech. Rep. 36.814 V9.0.0*, Mar 2010.

[19] G. Yuan, X. Zhang, W. Wang, and Y. Yang, "Carrier aggregation for LTE-advanced mobile communication systems," *IEEE Commun. Mag.*, vol. 48, pp. 88–93, Feb 2010.

[20] M. Iwamura, K. Etemad, M.-H. Fong, R. Nory, and R. Love, "Carrier aggregation framework in 3GPP LTE-advanced," *Communications Magazine, IEEE*, vol. 48, pp. 60–67, Aug 2010.

[21] K. Pedersen, T. Kolding, F. Frederiksen, I. Kovács, D. Laselva, and P. Mogensen, "An overview of downlink radio resource management for UTRAN long-term evolution," *Communications Magazine, IEEE*, vol. 47, pp. 86–93, Jul 2009.

[22] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); base station (BS) radio transmission and reception (release 9)," *Tech. Spec. 36.104 V9.4.0*, Jun 2010.

[23] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); physical channels and modulation (release 9)," *Tech. Spec. 36.211 V9.1.0*, Mar 2010.

[24] R. Knopp and P. Humblet, "Multiple-accessing over frequency-selective fading channels," in *Personal, Indoor and Mobile Radio Communications, 1995. PIMRC'95, Sixth IEEE International Symposium on*, vol. 3, pp. 1326–1330, Sep 1995.

[25] A. Goldsmith and S.-G. Chua, "Variable-rate variable-power MQAM for fading channels," *Communications, IEEE Transactions on*, vol. 45, pp. 1218–1230, Oct 1997.

[26] C. Y. Wong, R. Cheng, K. Lataief, and R. Murch, "Multiuser OFDM with adaptive subcarrier, bit, and power allocation," *Selected Areas in Communications, IEEE Journal on*, vol. 17, pp. 1747–1758, Oct 1999.

[27] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); physical layer procedures," *Tech. Spec. 36.213 V9.1.0*, Mar 2009.

[28] D. Kim and D. G. Jeong, "Capacity unbalance between uplink and downlink in spectrally overlaid narrow-band and wide-band CDMA mobile systems," *Vehicular Technology, IEEE Transactions on*, vol. 49, pp. 1086–1093, Jul 2000.

[29] W.-C. Wong, C.-E. Sundberg, and N. Seshadri, "Shared time division duplexing: an approach to low-delay high-quality wireless digital speech communications," *Vehicular Technology, IEEE Transactions on*, vol. 43, pp. 934–945, Nov 1994.

[30] Y. Kobayashi, K. Mori, and H. Kobayashi, "Radio resource assignment scheme for asymmetric traffic in CDMA/shared-TDD cellular packet communications," in *Vehicular Technology Conference, 2004. VTC2004-Fall. 2004 IEEE 60th*, vol. 2, pp. 939–943, Sep 2004.

[31] L. Chen, U. Yoshida, H. Murata, and S. Hirose, "Dynamic timeslot allocation algorithms suitable for asymmetric traffic in multimedia TDMA/TDD cellular radio," in *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, vol. 2, pp. 1424–1428, May 1998.

[32] P. W. C. Chan, E. S. Lo, R. R. Wang, E. K. S. Au, V. K. N. Lau, R. S. Cheng, W. H. Mow, R. D. Murch, and K. B. Letaief, "The evolution path of 4G networks: FDD or TDD?," *Communications Magazine, IEEE*, vol. 44, pp. 42–50, Dec 2006.

[33] Y. Zheng, "How to operate– boosting WiMAX indoor coverage," *Huawei Communicate*, pp. 31–32, Apr 2009.

[34] H. Claussen, L. Ho, and L. Samuel, "Financial analysis of a Pico-cellular home network deployment," in *Communications, 2007. ICC'07. IEEE International Conference on*, pp. 5604–5609, Jun 2007.

[35] B. Crow, I. Widjaja, L. Kim, and P. Sakai, "IEEE 802.11 wireless local area networks," *Communications Magazine, IEEE*, vol. 35, pp. 116–126, Sep 1997.

[36] S. ping Yeh, S. Talwar, S. choon Lee, and H. Kim, "WiMAX femtocells: a perspective on network architecture, capacity, and coverage," *Communications Magazine, IEEE*, vol. 46, pp. 58–65, Oct 2008.

[37] IEEE Instrumentation and Measurement Society, "IEEE standard for a precision clock synchronization protocol for networked measurement and control systems," *Std. 1588-2002*, Nov 2002.

[38] J. Mannermaa, K. Kalliomaki, T. Mansten, and S. Turunen, "Timing performance of various GPS receivers," in *Frequency and Time Forum, 1999 and the IEEE International Frequency Control Symposium, 1999., Proceedings of the 1999 Joint Meeting of the European*, vol. 1, pp. 287–290, Apr 1999.

[39] R. Leidenfrost and W. Elmenreich, "Firefly clock synchronization in an 802.15.4 wireless network," *EURASIP Journal on Embedded Systems*, vol. 2009, Mar 2009.

[40] R. Mangharam, A. Rowe, and R. Rajkumar, "FireFly: a cross-layer platform for real-time embedded wireless networks," *SpringerLink Journal on Real-Time Systems*, vol. 37, pp. 183–231, Dec 2007.

[41] IST-4-027756 WINNER II, "Final CG 'local area' description for integration into overall system concept and assessment of key technologies," *Tech. Rep. D6.13.12 V1.0*, Jan 2007.

[42] A. Tyrrell and G. Auer, "Imposing a reference timing onto firefly synchronization in wireless networks," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 222–226, Apr 2007.

[43] 3GPP, "Evolved universal terrestrial radio access (E-UTRA); TDD Home eNode B (HeNB) radio frequency (RF) requirements analysis," *Tech. Rep. 36.922 V9.1.0*, Jun 2010.

[44] A. Simonsson, "Frequency reuse and intercell interference co-ordination in E-UTRA," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 3091–3095, Apr 2007.

[45] R. Giuliano, C. Monti, and P. Loreti, "WiMAX fractional frequency reuse for rural environments," *Wireless Communications Magazine, IEEE*, vol. 15, pp. 60–65, Jun 2008.

[46] C. Wijting, K. Doppler, K. KallioJarvi, T. Svensson, M. Sternad, G. Auer, N. Johansson, J. Nystrom, M. Olsson, A. Osseiran, M. Dottling, J. Luo, T. Lestable, and S. Pfletschinger, "Key technologies for IMT-advanced mobile communication systems," *Wireless Communications Magazine, IEEE*, vol. 16, pp. 76–85, Jun 2009.

[47] WINNER+, "Initial report on system aspects of flexible spectrum use," *Tech. Rep. D1.2 V1.0*, Jan 2009.

[48] J.-P. Kermoal, S. Pfletschinger, K. Hooli, S. Thilakawardana, J. Lara, and Y. Zhu, "Spectrum sharing for WINNER radio access networks," in *Cognitive Radio Oriented Wireless Networks and Communications, 2006. 1st International Conference on*, pp. 1–5, Jun 2006.

[49] W.-Y. Lee and I. Akyildiz, "Joint spectrum and power allocation for inter-cell spectrum sharing in cognitive radio networks," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, pp. 1–12, Oct 2008.

[50] P. Lunden, E. Virtej, and A. Sorri, "Distributed flexible spectrum use for uncoordinated scenarios," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pp. 2871 –2875, Sep 2009.

[51] K. Johansson, J. Bergman, D. Gerstenberger, M. Blomgren, and A. Wallen, "Multi-carrier HSPA evolution," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pp. 1–5, Apr 2009.

[52] D. de Andrade, A. Klein, H. Holma, I. Viering, and G. Liebl, "Performance evaluation on dual-cell HSDPA operation," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, pp. 1–5, Sep 2009.

[53] T. Dean and P. Fleming, "Trunking efficiency in multi-carrier CDMA systems," in *Vehicular Technology Conference, 2002. Proceedings. VTC 2002-Fall. 2002 IEEE 56th*, vol. 1, pp. 156–160, Sep 2002.

[54] T. Wong and V. Prabhu, "Capacity growth for CDMA system: multiple sectors and multiple carriers deployment," in *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, vol. 2, pp. 1608–1611, May 1998.

[55] B. Song, J. C. Kim, and S. Oh, "Performance analysis of channel assignment methods for multiple carrier CDMA cellular systems," in *Vehicular Technology Conference, 1999 IEEE 49th*, vol. 1, pp. 10–14, Jul 1999.

[56] A. He, "Performance comparison of load balancing methods in multiple carrier CDMA systems," in *Personal, Indoor and Mobile Radio Communications, 2000. PIMRC 2000. The 11th IEEE International Symposium on*, vol. 1, pp. 113–118, 2000.

[57] N. Tripathi and S. Sharma, "Dynamic load balancing in a CDMA system with multiple carriers," in *Vehicular Technology Conference, 2001. VTC 2001 Fall. IEEE VTS 54th*, vol. 2, pp. 1010–1014, 2001.

[58] R. Attar, D. Ghosh, C. Lott, M. Fan, P. Black, R. Rezaiifar, and P. Agashe, "Evolution of CDMA2000 cellular networks: multicarrier EV-DO," *Communications Magazine, IEEE*, vol. 44, pp. 46–53, Mar 2006.

[59] N. Jindal, "MIMO broadcast channels with finite-rate feedback," *Information Theory, IEEE Transactions on*, vol. 52, pp. 5045–5060, Nov 2006.

[60] S. Sanayei and A. Nosratinia, "Opportunistic downlink transmission with limited feedback," *Information Theory, IEEE Transactions on*, vol. 53, pp. 4363–4372, Nov 2007.

[61] Y. Al-Harthi, A. Tewfik, and M.-S. Alouini, "Multiuser diversity with quantized feedback," *Wireless Communications, IEEE Transactions on*, vol. 6, pp. 330–337, Jan 2007.

[62] J. van de Beek, "Channel quality feedback schemes for 3GPP's Evolved-UTRA downlink," in *Global Telecommunications Conference, 2006. GLOBECOM'06. IEEE*, pp. 1–5, Dec 2006.

[63] P. Svedman, S. K. Wilson, L. J. Cimini, and B. Ottersten, "Opportunistic beamforming and scheduling for OFDMA systems," *Communications, IEEE Transactions on*, vol. 55, pp. 941–952, May 2007.

[64] N. Kolehmainen, J. Puttonen, P. Kela, T. Ristaniemi, T. Henttonen, and M. Moisio, "Channel quality indication reporting schemes for UTRAN long term evolution downlink," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 2522–2526, May 2008.

[65] Y. Sun, W. Xiao, R. Love, K. Stewart, A. Ghosh, R. Ratasuk, and B. Classon, "Multi-user scheduling for OFDM downlink with limited feedback for evolved UTRA," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, pp. 1–5, Sep 2006.

[66] Ericsson, "Uplink transmission of ACK/NACK for E-UTRA TDD," *3GPP T-doc R1-071894*, Apr 2007.

[67] M. Rahman and D. Astély, "Link level investigation of ACK/NACK bundling for LTE TDD," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pp. 1–5, Apr 2009.

[68] P. Mogensen, W. Na, I. Kovács, F. Frederiksen, A. Pokhariyal, K. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the Shannon bound," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 1234–1238, Apr 2007.

[69] J.-I. Chuang, "Performance limitations of TDD wireless personal communications with asynchronous radio ports," *Electronics Letters*, vol. 28, pp. 532–534, Mar 1992.

[70] H. Holma, S. Heikkinen, O.-A. Lehtinen, and A. Toskala, "Interference considerations for the time division duplex mode of the UMTS terrestrial radio access," *Selected Areas in Communications, IEEE Journal on*, vol. 18, pp. 1386–1393, Aug 2000.

[71] M. Morikura, S. Nitta, K. Kawazoe, H. Kazama, and S. Kato, "Semi-autonomous synchronization for TDMA-TDD communication system," in *Vehicular Technology Conference, 1993 IEEE 43rd*, pp. 297–300, May 1993.

[72] 3GPP, "NodeB synchronisation for TDD (release 4)," *Tech. Rep. 25.836 V4.1.0*, Mar 2001.

[73] M. Capaccioli and D. Rispo, "A technique to realize base stations on-air frame synchronization in TD-SCDMA radio system," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 2, pp. 982–986, Oct 2003.

[74] T. Wada, N. Saitou, T. Hara, and M. Okada, "Frame synchronization among base stations for TDD systems," in *Communications, Control and Signal Processing (ISCCSP), 2010 4th International Symposium on*, pp. 1–4, Mar 2010.

[75] R. Holler, T. Sauter, and N. Kero, "Embedded SynUTC and IEEE 1588 clock synchronization for industrial ethernet," in *Emerging Technologies and Factory Automation, 2003. Proceedings. ETFA'03. IEEE Conference*, vol. 1, pp. 422–426, Sep 2003.

[76] J. Kannisto, T. Vanhatupa, M. Hannikainen, and T. Hamalainen, "Software and hardware prototypes of the IEEE 1588 precision time protocol on wireless LAN," in *Local and Metropolitan Area Networks, 2005. LANMAN 2005. The 14th IEEE Workshop on*, Sep 2005.

[77] W. Jeong and M. Kavehrad, "Cochannel interference reduction in dynamic-TDD fixed wireless applications, using time slot allocation algorithms," *Communications, IEEE Transactions on*, vol. 50, pp. 1627–1636, Oct 2002.

[78] J. Frenzel, "Genetic algorithms," *Potentials, IEEE*, vol. 12, pp. 21–24, Oct 1993.

[79] S. Kumar, Y. Wang, N. Marchetti, I. Kovács, K. Pedersen, and P. Mogensen, "Spectrum load balancing for flexible spectrum usage in local area deployment scenario," in *New Frontiers in Dynamic Spectrum Access Networks, 2008. DySPAN 2008. 3rd IEEE Symposium on*, Oct 2008.

[80] L. Garcia, Y. Wang, S. Frattasi, N. Marchetti, P. Mogensen, and K. Pedersen, "Comparison of spectrum sharing techniques for IMT-A systems in local area networks," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, Apr 2009.

[81] S. Kumar, Y. Wang, and N. Marchetti, "Self-organized spectrum chunk selection algorithm for local area LTE-Advanced," in *ITU-T Kaleidoscope 2010*, Dec 2010.

[82] I. Katzela and M. Naghshineh, "Channel assignment schemes for cellular mobile telecommunication systems: a comprehensive survey," *Communications Surveys Tutorials, IEEE*, vol. 3, pp. 10–31, Second Quarter 2000.

[83] J. Wang, "Simulation and performance analysis of dynamic channel allocation algorithms in DECT," *Vehicular Technology, IEEE Transactions on*, vol. 42, pp. 563–569, Nov 1993.

[84] Y. Argyropoulos, S. Jordan, and S. Kumar, "Dynamic channel allocation in interference-limited cellular systems with uneven traffic distribution," *Vehicular Technology, IEEE Transactions on*, vol. 48, pp. 224–232, Jan 1999.

[85] L. Cimini and G. Foschini, "Distributed algorithms for dynamic channel allocation in microcellular systems," in *Vehicular Technology Conference, 1992, IEEE 42nd*, vol. 2, pp. 641–644, May 1992.

[86] S. Yasuda and S. Onoe, "Autonomous channel assignment control for flexible reuse in mobile radio systems," in *Vehicular Technology Conference, 1992, IEEE 42nd*, vol. 2, pp. 798–801, May 1992.

[87] D. Everitt and D. Manfield, "Performance analysis of cellular mobile communication systems with dynamic channel assignment," *Selected Areas in Communications, IEEE Journal on*, vol. 7, pp. 1172–1180, Oct 1989.

[88] S. Onoe and S. Yasuda, "Flexible re-use for dynamic channel assignment in mobile radio systems," in *Communications, 1989. ICC'89, BOSTON-ICC/89. Conference record. 'World Prosperity Through Communications', IEEE International Conference on*, vol. 1, pp. 472–476, Jun 1989.

[89] D. Cox and D. Reudink, "A comparison of some channel assignment strategies in large-scale mobile communications systems," *Communications, IEEE Transactions on*, vol. 20, pp. 190–195, Apr 1972.

[90] L. Garcia, K. Pedersen, and P. Mogensen, "Autonomous component carrier selection: interference management in local area environments for LTE-Advanced," *Communications Magazine, IEEE*, vol. 47, pp. 110–116, Sep 2009.

[91] Nokia Siemens Networks, "Use of background interference matrix for autonomous component carrier selection for LTE-Advanced," *3GPP T-doc R1-090235*, Feb 2009.

[92] A. Patel, J. Degesys, and R. Nagpal, "Desynchronization: the theory of self-organizing algorithms for round-robin scheduling," in *Self-Adaptive and Self-Organizing Systems, 2007. SASO'07. First International Conference on*, pp. 87–96, Jul 2007.

[93] J. Lim, H. Myung, K. Oh, and D. Goodman, "Channel-dependent scheduling of uplink single carrier FDMA systems," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, pp. 1–5, Sep 2006.

[94] J. Ramiro-Moreno, K. Pedersen, and P. Mogensen, "Network performance of transmit and receive antenna diversity in HSDPA under different packet scheduling strategies," in *Vehicular Technology Conference, 2003. VTC 2003-Spring. The 57th IEEE Semiannual*, vol. 2, pp. 1454–1458, Apr 2003.

[95] T. Kolding, "Link and system performance aspects of proportional fair scheduling in WCDMA/HSDPA," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 3, pp. 1717–1722, Oct 2003.

[96] C. Wengerter, J. Ohlhorst, and A. von Elbwart, "Fairness and throughput analysis for generalized proportional fair frequency scheduling in OFDMA," in *Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st*, vol. 3, pp. 1903–1907, May 2005.

[97] A. Pokhariyal, T. Kolding, and P. Mogensen, "Performance of downlink frequency domain packet scheduling for the UTRAN long term evolution," in *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pp. 1–5, Sep 2006.

[98] L. Chen, W. Chen, X. Zhang, and D. Yang, "Analysis and simulation for spectrum aggregation in LTE-Advanced system," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, Sep 2009.

[99] D. Zhang, P. Vitthaladevuni, B. Mohanty, and J. Hou, "Performance analysis of Dual-Carrier HSDPA," in *Vehicular Technology Conference (VTC 2010-Spring), 2010 IEEE 71st*, May 2010.

[100] China Mobile Communications Corporation (CMCC), "System simulation results on carrier aggregation for bursty traffic," *3GPP T-doc R1-091828*, May 2009.

[101] D. Knuth, *The Art of Computer Programming, Volume 3: Sorting and Searching*. Addison-Wesley, 1973.

[102] A. Pokhariyal, K. Pedersen, G. Monghal, I. Kovács, C. Rosa, T. Kolding, and P. Mogensen, "HARQ aware frequency domain packet scheduler with different degrees of fairness for the UTRAN long term evolution," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, pp. 2761–2765, Apr 2007.

[103] G. Monghal, K. Pedersen, I. Kovács, and P. Mogensen, "QoS oriented time and frequency domain packet schedulers for the UTRAN long term evolution," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 2532–2536, May 2008.

[104] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Vehicular Technology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 3, pp. 1854–1858, May 2000.

[105] H. Kushner and P. Whiting, "Convergence of proportional-fair sharing algorithms under general conditions," *Wireless Communications, IEEE Transactions on*, vol. 3, pp. 1250–1259, Jul 2004.

[106] J. Holtzman, "Asymptotic analysis of proportional fair algorithm," in *Personal, Indoor and Mobile Radio Communications, 2001 12th IEEE International Symposium on*, vol. 2, pp. F–33 –F–37, Sep 2001.

[107] M. Andrews, "A survey of scheduling theory in wireless data networks," *Wireless Communications: The IMA Volumes in Mathematics and its Applications*, vol. 143, pp. 1–17, 2007.

[108] Y. Wang, K. Pedersen, P. Mogensen, and T. Sørensen, "Carrier load balancing methods with bursty traffic for LTE-Advanced systems," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pp. 22–26, Sep 2009.

[109] K. Pedersen, T. Kolding, I. Kovács, G. Monghal, F. Frederiksen, and P. Mogensen, "Performance analysis of simple channel feedback schemes for a practical OFDMA system," *Vehicular Technology, IEEE Transactions on*, vol. 58, pp. 5309–5314, Nov 2009.

[110] P. Svedman, L. Cimini, and B. Ottersten, "Using unclaimed sub-carriers in opportunistic OFDMA systems," in *Vehicular Technology Conference, 2006. VTC-2006 Fall. 2006 IEEE 64th*, Sep 2006.

[111] J. Lee, J.-K. Han, and J. Zhang, "MIMO technologies in 3GPP LTE and LTE-Advanced," *EURASIP Journal on Wireless Communications and Networking*, May 2009.

[112] Y. Hu, D. Astély, R. Baldemair, and S. Falahati, "Semi-blind multi-user detection for LTE PUCCH," in *Wireless Communications and Networking Conference, 2009. WCNC 2009. IEEE*, pp. 1–5, Apr 2009.

[113] IST-4-027756 WINNER II, "WINNER II channel models," *Tech. Rep. D1.1.2*, Sep 2007.

[114] 3GPP, "Physical layer aspect for evolved universal terrestrial radio access (UTRA)," *Tech. Rep. 25.814 V7.1.0*, Sep 2006.

[115] S. Joshi, R. Cheung, P. Monajemi, and J. Villasenor, "Traffic-based study of Femtocell access policy impacts on HSPA service quality," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pp. 1–6, Dec 2009.

[116] J. Gora and T. Kolding, "Deployment aspects of 3G Femtocells," in *Personal, Indoor and Mobile Radio Communications, 2009 IEEE 20th International Symposium on*, pp. 1507–1511, Sep 2009.

[117] L. G. U. Garcia, K. I. Pedersen, and P. E. Mogensen, "On open versus closed LTE-Advanced Femtocells and dynamic interference coordination," in *Wireless Communications and Networking Conference (WCNC), 2010 IEEE*, Apr 2010.

[118] S. Sesia, B. Matthew, and T. Issam, *LTE, the UMTS long term evolution: from theory to practice*. John Wiley & Sons, Feb 2009.

[119] A. Simonsson and A. Furuskar, "Uplink power control in LTE - overview and performance," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pp. 1–5, Sep 2008.

[120] C. Castellanos, D. Villa, C. Rosa, K. Pedersen, F. Calabrese, P.-H. Michaelsen, and J. Michel, "Performance of uplink fractional power control in UTRAN LTE," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 2517–2521, May 2008.

[121] B. Muhammad and A. Mohammed, "Performance evaluation of uplink closed loop power control for LTE system," in *Vehicular Technology Conference Fall (VTC 2009-Fall), 2009 IEEE 70th*, Sep 2009.

[122] M. Boussif, C. Rosa, J. Wigard, and R. Mullner, "Load adaptive power control in LTE uplink," in *Wireless Conference (EW), 2010 European*, pp. 288–293, Feb 2010.

[123] M. Boussif, N. Quintero, F. Calabrese, C. Rosa, and J. Wigard, "Interference based power control performance in LTE uplink," in *Wireless Communication Systems. 2008. ISWCS'08. IEEE International Symposium on*, pp. 698–702, Oct 2008.

[124] L. Garcia, K. Pedersen, and P. Mogensen, "Uplink performance of dynamic interference coordination under fractional power control for LTE-Advanced femtocells," in *Vehicular Technology Conference, 2010. VTC2010-Fall. IEEE 72nd*, Sep 2010.

[125] 3GPP, "Radio frequency (RF) system scenarios (release 9)," *Tech. Rep. 25.942 V9.0.0*, Dec 2009.

[126] 3GPP, "Deployment scenarios (release 9)," *Tech. Rep. 25.943 V9.0.0*, Dec 2009.

[127] K. Brueninghaus, D. Astély, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Personal, Indoor and Mobile Radio Communications, 2005. PIMRC 2005. IEEE 16th International Symposium on*, vol. 4, pp. 2306–2311, Sep 2005.

[128] A. Sampath, P. Sarath Kumar, and J. Holtzman, "On setting reverse link target SIR in a CDMA system," in *Vehicular Technology Conference, 1997 IEEE 47th*, vol. 2, pp. 929–933, May 1997.

[129] J. Winters, "Optimum combining in digital mobile radio with cochannel interference," *Vehicular Technology, IEEE Transactions on*, vol. 33, pp. 144–155, Aug 1984.

[130] A. Barreto and R. Vieira, "A critical analysis of receiver diversity with multiple antennas with spatially coloured noise," in *Signal Processing Advances in Wireless Communications, 2008. SPAWC 2008. IEEE 9th Workshop on*, pp. 471–475, Jul 2008.

[131] 3GPP, "Text proposal for evaluation methodology," *3GPP T-doc R1-084026*, Sep 2008.

[132] G. Foschini, "Layered space-time architecture for wireless communications in a fading environment when using multi-element antennas," *Bell Labs Technical Journal*, vol. 1, pp. 41–59, Autumn 1996.

[133] L. Zheng and D. Tse, "Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels," *Information Theory, IEEE Transactions on*, vol. 49, pp. 1073–1096, May 2003.

[134] K. Etemad, "Overview of mobile WiMAX technology and evolution," *Communications Magazine, IEEE*, vol. 46, pp. 31–40, Oct 2008.

[135] A. Goldsmith and P. Varaiya, "Capacity of fading channels with channel side information," *Information Theory, IEEE Transactions on*, vol. 43, pp. 1986–1992, Nov 1997.

[136] M.-S. Alouini and A. Goldsmith, "Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques," *Vehicular Technology, IEEE Transactions on*, vol. 48, pp. 1165–1181, Jul 1999.

[137] S. Catreux, P. Driessen, and L. Greenstein, "Data throughputs using multiple-input multiple-output (MIMO) techniques in a noise-limited cellular environment," *Wireless Communications, IEEE Transactions on*, vol. 1, pp. 226–235, Apr 2002.

[138] P. Smith and M. Shafi, "On a Gaussian approximation to the capacity of wireless MIMO systems," in *Communications, 2002. ICC 2002. IEEE International Conference on*, vol. 1, pp. 406–410, Apr 2002.

[139] E. Liu and K. Leung, "Proportional fair scheduling: Analytical insight under Rayleigh fading environment," in *Wireless Communications and Networking Conference, 2008. WCNC 2008. IEEE*, pp. 1883–1888, Mar 2008.

[140] J.-G. Choi and S. Bahk, "Cell-throughput analysis of the proportional fair scheduler in the single-cell environment," *Vehicular Technology, IEEE Transactions on*, vol. 56, pp. 766–778, Mar 2007.

[141] Z. Han, Z. Ji, and K. Liu, "Fair multiuser channel allocation for OFDMA networks using Nash bargaining solutions and coalitions," *Communications, IEEE Transactions on*, vol. 53, pp. 1366–1376, Aug 2005.

[142] S. Catreux, P. Driessen, and L. Greenstein, "Simulation results for an interference-limited multiple-input multiple-output cellular system," *Communications Letters, IEEE*, vol. 4, pp. 334–336, Nov 2000.

[143] P. Kela, J. Puttonen, N. Kolehmainen, T. Ristaniemi, T. Henttonen, and M. Moisio, "Dynamic packet scheduling performance in UTRA Long Term Evolution downlink," in *Wireless Pervasive Computing, 2008. ISWPC 2008. 3rd International Symposium on*, pp. 308–313, May 2008.

[144] N. Wei, A. Pokhariyal, T. Sørensen, T. Kolding, and P. Mogensen, "Performance of spatial division multiplexing MIMO with frequency domain packet scheduling: From theory to practice," *Selected Areas in Communications, IEEE Journal on*, vol. 26, pp. 890–900, Aug 2008.

[145] W. Jakes, *Microwave mobile communications*. John Wiley & Sons, 1974.

[146] S. Alamouti, "A simple transmit diversity technique for wireless communications," *Selected Areas in Communications, IEEE Journal on*, vol. 16, pp. 1451–1458, Oct 1998.

[147] I. Kovács, K. Pedersen, J. Wigard, F. Frederiksen, and T. Kolding, "HSDPA performance in mixed outdoor-indoor Micro cell scenarios," in *Personal, Indoor and Mobile Radio Communications, 2006 IEEE 17th International Symposium on*, pp. 1–5, Sep 2006.

[148] K. Leonard, *Queueing Systems Volume 1: Theory*. John Wiley & Sons, 1975.

[149] D. Hong and S. Rappaport, "Traffic model and performance analysis for cellular mobile radio telephone systems with prioritized and nonprioritized handoff procedures," *Vehicular Technology, IEEE Transactions on*, vol. 35, pp. 77–92, Aug 1986.

[150] Y. Liang, Y. Wu, R. Liu, and C. Chen, "Adaptive decision method for the Best M reporting scheme," in *Wireless, Mobile and Sensor Networks, 2007. (CCWMSN07). IET Conference on*, pp. 340–343, Dec 2007.

[151] Sharp, "Best-M based scheme for CQI reporting in the LTE UL," *3GPP T-doc R1-073320*, Aug 2007.

[152] M. Sternad, T. Ottosson, A. Ahlen, and A. Svensson, "Attaining both coverage and high spectral efficiency with adaptive OFDM downlinks," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 4, pp. 2486–2490, Sep 2003.

[153] T. Rappaport and R. Brickhouse, "A simulation study of urban in-building cellular frequency reuse," *Personal Communications, IEEE*, vol. 4, pp. 19–23, Feb 1997.

[154] WiMAX forum, "Mobile WiMAX - Part I: a technical overview and performance evaluation," Aug 2006.

[155] J. Tomcik, "MBFDD and MBTDD wideband mode: technology overview," *IEEE 802.20-05/68r1*, Jan 2006.

[156] G. Boudreau, J. Panicker, N. Guo, R. Chang, N. Wang, and S. Vrzic, "Interference coordination and cancellation for 4G networks," *Communications Magazine, IEEE*, vol. 47, pp. 74–81, Apr 2009.

[157] A. Stolyar and H. Viswanathan, "Self-organizing dynamic fractional frequency reuse in OFDMA systems," in *INFOCOM 2008. The 27th Conference on Computer Communications. IEEE*, pp. 691–699, Apr 2008.

[158] H. Fujii and H. Yoshino, "Theoretical capacity and outage rate of OFDMA cellular system with fractional frequency reuse," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 1676–1680, May 2008.

[159] Y. Zhou and N. Zein, "Simulation study of fractional frequency reuse for mobile WiMAX," in *Vehicular Technology Conference, 2008. VTC Spring 2008. IEEE*, pp. 2592–2595, May 2008.

[160] M. Assaad, "Optimal fractional frequency reuse (FFR) in multicellular OFDMA system," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pp. 1–5, Sep 2008.

[161] H. Luo, Z. Zhang, H. Jia, G. Yu, and S. Li, "Performance comparison of IEEE 802.16e and IEEE 802.20 systems under different frequency reuse schemes," in *Vehicular Technology Conference, 2008. VTC 2008-Fall. IEEE 68th*, pp. 1–5, Sep 2008.

[162] D. Lopez-Perez, A. Juttner, and J. Zhang, "Dynamic frequency planning versus frequency reuse schemes in OFDMA networks," in *Vehicular Technology Conference, 2009. VTC Spring 2009. IEEE 69th*, pp. 1–5, Apr 2009.